



HAL
open science

Intrinsic Bayesian inference on a Poisson rate and on the ratio of two Poisson rates

Stéphane Laurent

► **To cite this version:**

Stéphane Laurent. Intrinsic Bayesian inference on a Poisson rate and on the ratio of two Poisson rates. *Journal of Statistical Planning and Inference*, 2012, doi:10.1016/j.jspi.2012.02.040. 10.1016/j.jspi.2012.02.040 . hal-00505234v3

HAL Id: hal-00505234

<https://hal.science/hal-00505234v3>

Submitted on 1 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Intrinsic Bayesian inference on a Poisson rate and on the ratio of two Poisson rates

Stéphane Laurent

*Université de Strasbourg & CNRS
IRMA, UMR 7501, 7 rue René Descartes
67084 Strasbourg Cedex, France*

Abstract

Our purpose is to explore the intrinsic Bayesian inference on the rate of a Poisson distribution and on the ratio of the rates of two independent Poisson distributions, with the natural conjugate family of priors in the first case and the semi-conjugate family of priors defined by Laurent & Legrand (2011) in the second case. Intrinsic Bayesian inference is derived from the Bayesian decision theory framework based on the intrinsic discrepancy loss function. We cover in particular the case of some objective Bayesian procedures suggested by Bernardo when considering reference priors.

Keywords: Poisson rates; Reference priors; Intrinsic loss; Intrinsic Discrepancy; Hypothesis testing; Semi-conjugate family

1. Introduction

Loss functions lie in the heart of statistical decision theory, which is commonly recognized as a coherent foundational framework for inferential problems and comparison of procedures. But, similarly to the choice of a prior distribution in a Bayesian approach, the choice of the loss function comes under criticism because of its arbitrariness. *Noninformative priors* have been proposed as candidates for an ‘objective’ choice of a prior distribution, and similarly, *intrinsic loss functions* (or, for short, intrinsic losses) have been proposed as candidates for an ‘objective’ choice of a loss function. As for a noninformative prior, an intrinsic loss function is directly derived from and only from the considered parametric statistical model. Robert (1996) provides a short and noteworthy discussion and main references on these two topics.

Considering a statistical model $\{p_\theta\}$ with parameter θ , an intrinsic loss $\ell(\theta_0, \theta)$ is *parametrization-free*: it depends on θ and θ_0 only through the two sampling distributions p_{θ_0} and p_θ . Then $\ell(\theta_0, \theta)$ measures how different p_{θ_0} and p_θ are, whereas usual metric losses such as the squared distance $(\theta_0 - \theta)^2$ or the absolute deviation $|\theta_0 - \theta|$ measure how different the parameters θ_0 and θ are. For example, when $\theta = \theta_0 + 10$, it is intuitively sound to require that the

difference between the two Poisson distributions $\mathcal{P}(\theta)$ and $\mathcal{P}(\theta_0)$ should not be as pronounced for large values of θ_0 than for small ones.

Robert (1996) studies some elementary properties of two intrinsic losses for the problem of evaluating estimators, namely the Hellinger loss and the entropy loss. A slight variant of the entropy loss, the *intrinsic discrepancy*, has been introduced by Bernardo and Rueda (2002), and further explored by Bernardo and Juarez (2003). The intrinsic discrepancy loss may also be used for the problem of Bayesian hypothesis testing (Bernardo 1999, Bernardo and Rueda 2002, Bernardo 2005a, Bernardo 2010), and the problem of producing Bayesian credible regions (Bernardo 2005b, Bernardo 2010).

In this paper, we mainly explore *intrinsic hypothesis testing* and *intrinsic credible intervals* for the rate of a Poisson distribution (in section 2) and the ratio of the rates of two independent Poisson distributions (in section 3), which respectively are hypothesis testing and credible regions derived from the Bayesian decision theory based on the intrinsic discrepancy loss. There is no theoretical development in this paper, but only some applications and numerical studies of the intrinsic inference theory to these two models.

We consider in particular the case of the *reference prior*, hence our work provides the *Bayesian reference criterion* for hypothesis testing, and the *reference intrinsic credible regions*, according to the terminology of Bernardo. We refer to Bernardo (2005a) for a survey on reference priors, which are the most successful and the most popular available noninformative priors, and we refer to Bernardo (2005a, 2009, 2010) for details on intrinsic inference. More generally, we consider the natural conjugate family of priors for the ‘one Poisson sample model’, and the semi-conjugate family of priors introduced by Laurent and Legrand (2011) for the ‘two Poisson samples model’.

Our main concluding observations are the following. Concerning the intrinsic credible intervals, whether we use the reference prior or a subjective prior, these ones are quite similar to the commonly used HPD and equal-tailed credible intervals. But the intrinsic credible intervals provide something more: they are based on the posterior expected intrinsic loss (for short, the *posterior loss*), which offers a measure of compatibility of the possible values of the unknown parameter with the observed data in terms of the information theoretic notion of divergence; we will see that two different situations can yield the same posterior distribution, but two very different posterior losses. Hence, two Bayesians performing an intrinsic inference can draw different conclusions although they get the same posterior distribution. Concerning the intrinsic hypothesis testing, this one is based on a posterior loss that measures the evidence against the null hypothesis, therefore it provides something more than only the posterior probability of this hypothesis. When using the reference prior, we will see that the decisions of the intrinsic hypothesis test are quite similar to those of the customary test based on the posterior probability of the null hypothesis, for an appropriate choice of the critical values used to define the rejection regions of these two tests. On the contrary, we will see that these two tests can yield very different conclusions when using a subjective prior. It will be striking that the intrinsic hypothesis test takes the sample size into account in addition to the

posterior distribution, and hence it seems more sensible than the test based on the posterior probability of the null hypothesis. We will also see that its behavior is more attractive than the one of the Bayesian test based on the Bayes factor.

2. One Poisson rate with conjugate family

We consider the statistical model given by one observation x assumed to be generated from a Poisson distribution $\mathcal{P}(\mu\tau)$ with known ‘observation-opportunity size’ τ , or ‘sample size’, and unknown rate parameter μ . We denote by

$$\delta(\mu \mid \mu_0) = E \left(\log \frac{\mathcal{P}(x \mid \mu\tau)}{\mathcal{P}(x \mid \mu_0\tau)} \mid \mu, \phi \right)$$

the Kullback-Leibler divergence from $\mathcal{P}(\mu\tau)$ to $\mathcal{P}(\mu_0\tau)$, and we denote by

$$\ell(\mu_0, \mu) = \min \{ \delta(\mu \mid \mu_0), \delta(\mu_0 \mid \mu) \}$$

the *intrinsic discrepancy* between $\mathcal{P}(\mu\tau)$ and $\mathcal{P}(\mu_0\tau)$. The unit of the intrinsic discrepancy would be the *bit* of information if the logarithm had been taken in base 2; instead of base e , as we do, in which case the unit is the *nat* of information.

In subsection 2.1, we derive $\ell(\mu_0, \mu)$ and the posterior expected intrinsic discrepancy

$$\bar{\ell}(\mu_0 \mid x) = \int \ell(\mu_0, \mu) \pi(\mu \mid x) d\mu \quad (2.1)$$

when the prior $\pi(\mu)$ on μ belongs to the conjugate family of Gamma distributions $\mathcal{G}(a, b)$ with shape parameter $a > 0$ and rate parameter $b \geq 0$. This covers in particular the case of the reference prior which is the improper Gamma distribution $\mathcal{G}(a, b)$ with $a = 0.5$ and $b = 0$ (see Bernardo & Smith, 1994).

In subsection 2.2, we study the lowest posterior loss credible intervals for μ based on the intrinsic discrepancy loss. In subsection 2.3 we study the Bayes hypothesis testing for $H_0: \{\mu \in M_0\}$ based on the loss $\ell(M_0, \mu)$ between M_0 and μ defined by $\ell(M_0, \mu) = \inf_{\mu_0 \in M_0} \ell(\mu_0, \mu)$.

2.1. Intrinsic discrepancy loss and posterior loss

By an easy calculation, or by an application of formula (A.2), one gets

$$\delta(\mu \mid \mu_0) = \tau h(\mu \mid \mu_0)$$

where

$$h(\mu \mid \mu_0) = \mu_0 - \mu + \mu \log \frac{\mu}{\mu_0}.$$

By elementary analysis, we get that the intrinsic discrepancy is given by

$$\ell(\mu_0, \mu) = \tau h(\min\{\mu_0, \mu\} \mid \max\{\mu_0, \mu\})$$

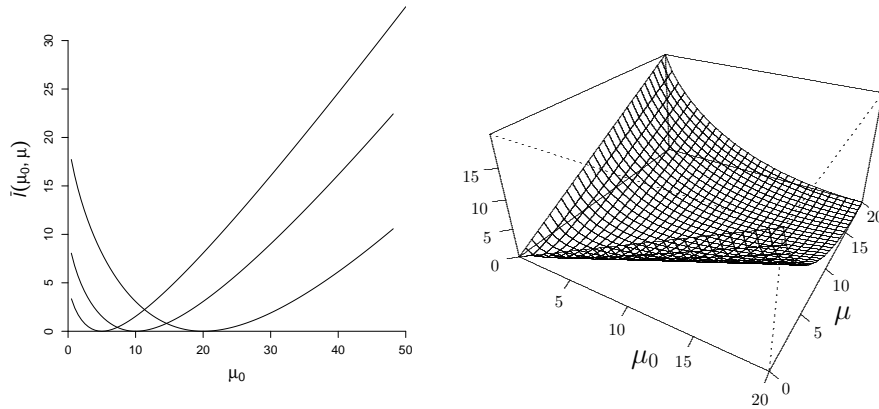


Figure 1: *Left:* $\mu_0 \mapsto \ell(\mu_0, \mu)$ for $\tau = 1$ and $\mu = 5, 10, 20$ (the corresponding curve vanishes at $\mu_0 = \mu$). *Right:* $(\mu_0, \mu) \mapsto \ell(\mu_0, \mu)$ with $\tau = 1$.

and one easily checks that $\mu_0 \mapsto \ell(\mu_0, \mu)$ is a continuous convex function. Figure 1 shows the curves $\mu_0 \mapsto \ell(\mu_0, \mu)$ for $\tau = 1$ and $\mu \in \{5, 10, 20\}$, and the surface $(\mu_0, \mu) \mapsto \ell(\mu_0, \mu)$ for $\tau = 1$.

Using the Gamma prior distribution $\mathcal{G}(a, b)$ on μ , the posterior expected intrinsic discrepancy (for short, the posterior loss) defined by (2.1) is

$$\begin{aligned} \bar{\ell}(\mu_0 | x) = \tau \left(\int_0^{\mu_0} h(\mu | \mu_0) \mathcal{G}(\mu | a + x, b + \tau) d\mu \right. \\ \left. + \int_{\mu_0}^{\infty} h(\mu_0 | \mu) \mathcal{G}(\mu | a + x, b + \tau) d\mu \right). \end{aligned} \quad (2.2)$$

The posterior loss $\bar{\ell}(\mu_0 | x)$ is a positive statistic that measures the compatibility of μ_0 with the observed data x . Note that $\bar{\ell}(\mu_0 | x)$ is a continuous convex function of μ_0 ; this is inherited from the same properties for $\ell(\mu_0, \mu)$.

2.2. Intrinsic credible intervals

The intrinsic credible intervals are more clearly understood after having introduced the *intrinsic estimate* $\hat{\mu}$ of μ . This one is defined as the Bayes estimate μ associated to the intrinsic discrepancy loss, that is, $\hat{\mu}$ is the value of μ_0 minimizing the posterior expected intrinsic discrepancy $\bar{\ell}(\mu_0 | x)$. An appealing property of the intrinsic estimate is its invariance under reparameterization; for example the intrinsic estimate of $\log(\mu)$ is nothing but $\log(\hat{\mu})$. This obviously results from the fact that the intrinsic discrepancy loss is parameterization-free.

By moving from this point estimation procedure to an interval estimation procedure, Bernardo (2005b) defined *intrinsic credible regions* as the lowest posterior loss credible regions when using the intrinsic discrepancy as the loss function. Thus, the *intrinsic 100q%-credible interval* I_x for μ is given by

$$I_x = \{\mu_0 | \bar{\ell}(\mu_0 | x) < c(x)\}$$

where $c(x)$ is chosen in order that $\pi(\mu \in I_x | x) = q$. This is indeed an interval by virtue of the shape of the posterior loss. It is easy to compute the intrinsic credible interval when the inverse cumulative posterior distribution of μ is available, in a way similar to that used to compute the HPD credible interval.

Actually our terminology differs from that of Bernardo (2005b): we allow an arbitrary choice of the prior, whereas Bernardo (2005b) always requires the use of the reference prior (the case of $a = 0.5$ and $b = 0$ in our context); in that case, we will say that $\hat{\mu}$ is the *reference intrinsic estimator*, I_x is the *reference intrinsic 100q%-credible interval*, and $\bar{\ell}(\mu_0 | x)$ is the *reference intrinsic posterior loss*. This is also the terminology adopted by Bernardo (2010).

Figure 2 shows an example of the reference intrinsic posterior loss $\bar{\ell}(\mu_0 | x)$ along with the reference intrinsic 95%-credible interval when $\tau = 50$, $x = 10$ and $x = 0$.

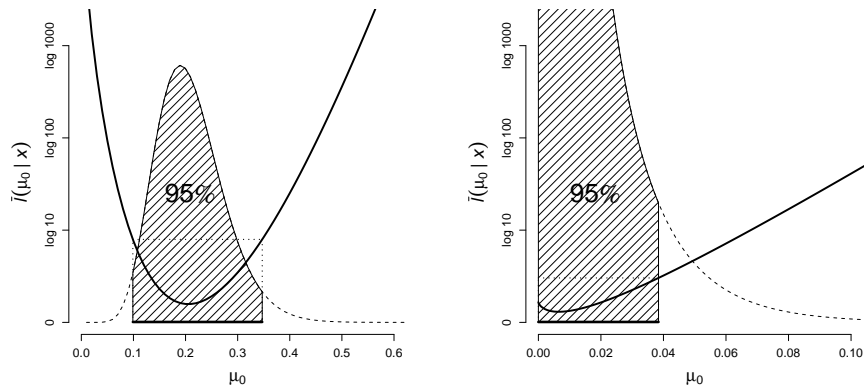


Figure 2: Reference posterior loss (solid), reference posterior distribution (dashed), and reference intrinsic 95%-credible interval. *Left:* $\tau = 50$ and $x = 10$. *Right:* $\tau = 50$ and $x = 0$.

Similarly to the intrinsic estimator, one appealing property of the intrinsic credible intervals I_x is its invariance under reparameterization; that is, for example, the intrinsic 100q%-credible interval for $\log(\mu)$ is simply obtained by applying the log function to the bounds of I_x . More particularly, the reference intrinsic credible interval enjoys the property of invariance under reparameterization from a higher point of view: the procedure consisting in deriving the reference prior and then the reference intrinsic credible interval is itself invariant under reparameterization, because of the invariance property under reparameterization for reference priors (see Bernardo 2005a). The invariance property also holds for the equal-tailed (probability centered) credible interval but on the contrary it fails for the HPD credible interval. Intrinsic credible intervals also enjoy an advantage as compared to equal-tailed credible intervals in situations such as the one on the right panel of figure 2: in this situation, the most likely values of μ under the posterior distribution are in the neighborhood of zero, hence the equal-tailed credible interval is not appropriate because it does not

contain these values.

For now we will do some investigations on intrinsic credible intervals when considering the case of the reference prior and the case of informative priors.

Case of the reference prior. In case of the reference prior ($a = 0.5, b = 0$), we have checked on numerous examples that the intrinsic $100q\%$ -credible intervals are close to the HPD and equal-tailed $100q\%$ -credible intervals. These intervals are plotted on figure 3 when $q = 0.95, \tau = 50$ and $x \in \{0, \dots, 15\}$.

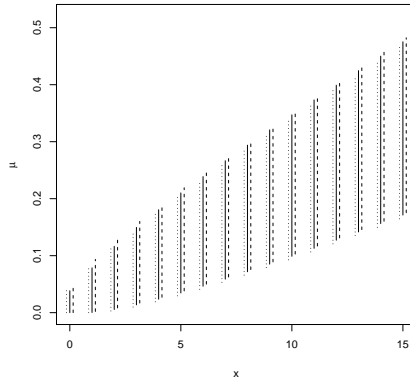


Figure 3: 95%-credible intervals with the reference prior, $\tau = 50$ and $x \in \{0, \dots, 15\}$. *Dotted*: HPD intervals. *Solid*: intrinsic intervals. *Dashed*: equal-tailed intervals.

The frequentist coverage $\mu \mapsto P(I_x \ni \mu \mid \mu)$ is plotted on figure 4 when I_x is the intrinsic, HPD, and equal-tailed 95%-credible interval, for $\tau = 30$ and $\tau = 100$. It tends to be close to 95% for each interval as long as the values of τ and μ are not too small. This so-called “*frequentist-matching*” property of credible intervals based on noninformative priors, which says that the frequentist coverage of these intervals is close to their posterior probability, has been considerably investigated for the equal-tailed and HPD intervals. We refer to Datta & Sweeting (2005) and references therein for details on this topic.

Case of an informative prior. We will see that intrinsic credible intervals still remain close to the HPD and equal-tailed credible intervals when using an arbitrary prior from the conjugate Gamma family. Let us call the four-tuple (τ, a, b, x) a *scenario*. Although the posterior loss $\bar{\ell}(\mu_0 \mid x)$, as we see from (2.2), depends on the scenario (τ, a, b, x) not only through the posterior distribution $\pi(\mu \mid x)$ but also through the sample size τ , the intrinsic $100q\%$ -credible interval depends on the scenario only through $\pi(\mu \mid x)$, since we can easily see from (2.2) that two scenarios with the same posterior distribution yield proportional posterior losses. On the contrary we will see in section 2.3 that the *intrinsic hypothesis test* yields results depending not only of the posterior distribution but also of the sample size.

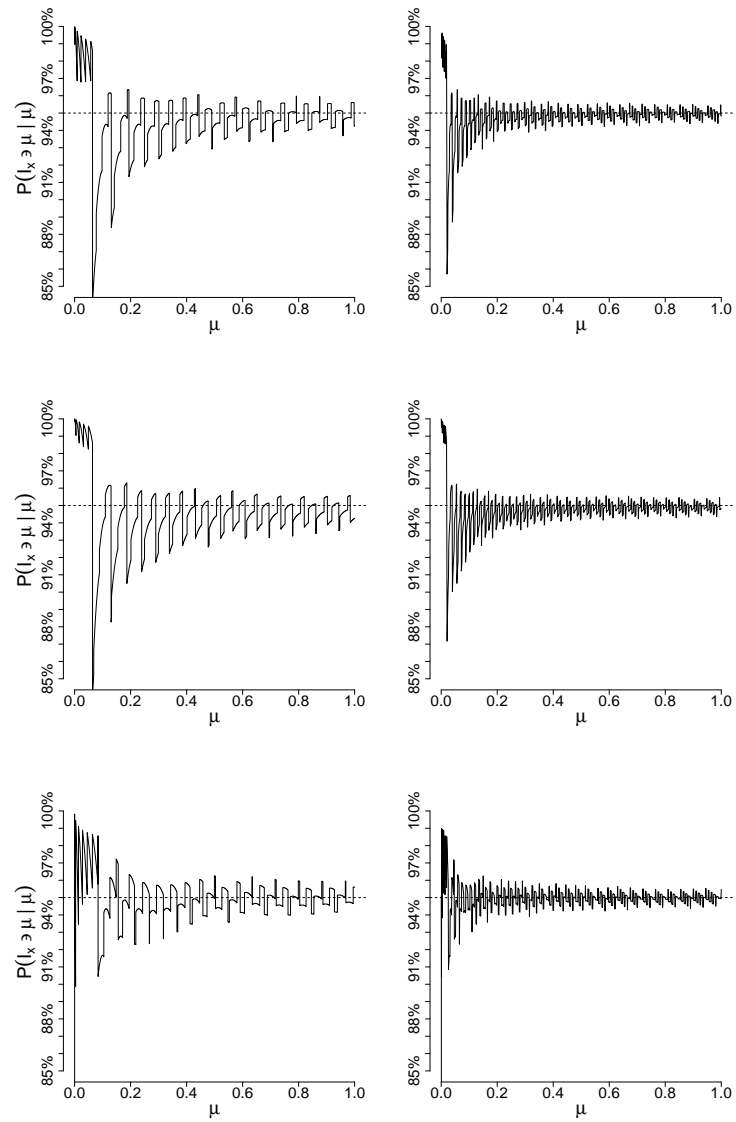


Figure 4: Frequentist coverage of 95%-credible intervals with the reference prior. *Left:* $\tau = 30$. *Right:* $\tau = 100$. *Top:* intrinsic interval. *Middle:* HPD interval. *Bottom:* equal-tailed interval.

Consider for instance the two scenarios of figure 5:

- $\tau = 300$, we use the reference prior ($a = 0.5$, $b = 0$) and we observe $x = 116$ cases (first scenario);
- $\tau = 50$, we use the $\mathcal{G}(a, b)$ prior with $a = 100.5$ and $b = 250$ and we observe $x = 16$ cases (second scenario).

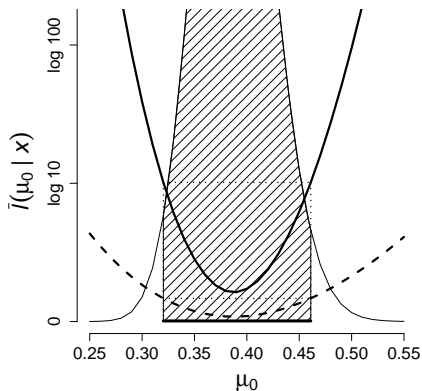


Figure 5: Posterior density (thin solid curve) on μ for both scenarios, posterior loss for each scenario (*first scenario*: bold solid curve; *second scenario*: bold dashed curve), and intrinsic 95%-credible interval for both scenarios.

In both cases, the posterior distribution on μ is $\mathcal{G}(116.5, 300)$, and, by (2.2), the posterior loss of the first scenario equals $300/50 = 6$ times the posterior loss of the second scenario; figure 5 shows the posterior loss for each scenario along with the shared posterior distribution and intrinsic 95%-credible interval. As a consequence, the intrinsic credible interval still remains close to the HPD and the equal-tailed credible intervals when one uses an informative prior, since we have previously seen (figure 3) that this holds for the reference prior, and since any scenario yields the same posterior distribution as a certain scenario with the reference prior, owing to the conjugacy property of the Gamma family.

However we may ask ourselves about the following fact we observe on figure 5: in the first scenario, with the noninformative prior and $\tau = 300$, the values of μ_0 outside the intrinsic credible interval are those for which the posterior loss $\bar{l}(\mu_0 | x)$ is, approximately, greater than $\log 10$, whereas in the second scenario, with the informative prior and $\tau = 50$, the values of μ_0 outside the intrinsic credible interval are, approximately, those for which $\bar{l}(\mu_0 | x) > \log 1.47$. Thus, if we agree that a value of $\bar{l}(\mu_0 | x)$ lower than $\log 10$ does not indicate a significant evidence against the likeliness of μ_0 (analogously to the conventions proposed for the intrinsic hypothesis test in subsection 2.3), we could be tempted to conclude that the intrinsic credible interval leaves out a lot of values which are not significantly unlikely. But one can argue that it is sensible to be less severe

when using an informative prior, because this one includes a part of available information on μ .

Lower posterior fixed loss interval, and precise hypothesis testing.

The intrinsic 100 q %-credible interval I_x consists of those values of μ for which the posterior loss $\bar{\ell}(\mu | x)$ does not exceed its posterior 100 q %-quantile. Observe on figure 2 that the higher value of $\bar{\ell}(\mu_0 | x)$ for $\mu_0 \in I_x$ significantly differs in the two considered cases. By interpreting $\bar{\ell}(\mu_0 | x)$ as a measure of evidence against μ_0 , this is rather sensible in light of the following comment: a small value of x indicates likeliness towards small values of μ , and, as suggested by the Fisher information of the model (proportional to $\mu \mapsto \mu^{-1}$), small values of μ should be more easily discriminated than higher ones.

We now have a look at the “confidence” interval

$$J_x = \{\mu_0 \mid \bar{\ell}(\mu_0 | x) < K\}$$

consisting of those values of μ_0 for which the posterior loss does not exceed a fixed value K . Of course it enjoys the invariance property under reparameterization but it is not a credible interval. Note that J_x is the confidence interval corresponding to the family of two-sided tests T_{μ_0} , $\mu_0 > 0$, where T_{μ_0} rejects the null hypothesis $H_0: \{\mu = \mu_0\}$ when $\bar{\ell}(\mu_0 | x) \geq K$. Actually, with the terminology of Bernardo (2010), the test T_{μ_0} is the *Bayes test criterion* for H_0 associated to the intrinsic discrepancy loss, and more particularly is the *reference test criterion* when one uses the reference prior, or the *Bayesian reference criterion* with the terminology of Bernardo (2005a). This is the particular case of *intrinsic hypothesis testing* for a null hypothesis of the form $H_0: \{\mu \in M_0\}$ when the set M_0 is the singleton $\{\mu_0\}$.

We restrict our attention to the case of the reference prior ($a = 0.5, b = 0$). Figure 6 shows the frequentist coverage $\mu \mapsto P(J_x \ni \mu \mid \mu)$ for $\tau = 30$ and $\tau = 100$ when $K = \log 10$ and $K = \log 100$. We see that the frequentist coverage has a decreasing trend and tends to stabilize above a certain level when μ increases. Note that the significance level of the test T_{μ_0} introduced above is provided by figure 6 since it equals $1 - P(J_x \ni \mu_0 \mid \mu_0)$.

2.3. Intrinsic hypothesis testing for one-sided hypotheses

Given $M_0 \subset (0, +\infty)$, the *intrinsic hypothesis test* for $H_0: \{\mu \in M_0\}$ is derived with the help of the intrinsic discrepancy $\ell(M_0, \mu)$ from M_0 to μ which is defined by $\ell(M_0, \mu) = \inf_{\mu_0 \in M_0} \ell(\mu_0, \mu)$, analogously to a distance between a point and a set. We refer to Bernardo (2005a, 2009, 2010) and references therein for more details. Again, our terminology differs from the one of Bernardo (2005a): we say, as Bernardo (2010) does, that the intrinsic hypothesis test in case of the reference prior is the *reference intrinsic hypothesis test*, whereas Bernardo (2005a) simply call it the *intrinsic hypothesis test*. The rejection rule of the intrinsic hypothesis test (also called the *Bayesian reference criterion* in case of the reference prior) consists in rejecting H_0 if $\bar{\ell}(M_0 | x) > \ell^*$ where ℓ^* is

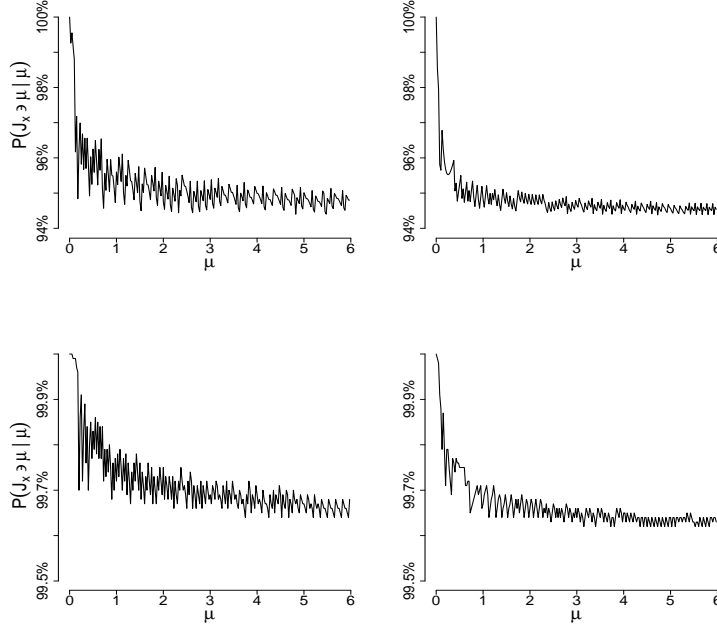


Figure 6: Frequentist coverage of J_x when using the reference prior. *Left:* $\tau = 30$. *Right:* $\tau = 100$. *Top:* $K = \log 10$. *Bottom:* $K = \log 100$.

a given utility constant and

$$\bar{\ell}(M_0 | x) = \int \ell(M_0, \mu) \pi(\mu | x) d\mu$$

is the posterior expected value of the intrinsic discrepancy from M_0 to μ , briefly termed hereafter as the posterior loss.

We focus on the case when $M_0 = (\mu^*; \infty)$ for a given $\mu^* > 0$, for which

$$\ell(M_0, \mu) = \begin{cases} 0 & \text{if } \mu \geq \mu^*, \\ \ell(\mu^*, \mu) & \text{if } \mu < \mu^*, \end{cases}$$

because $\ell(\mu_0, \mu_0) = 0$ and $\mu_0 \mapsto \ell(\mu_0, \mu)$ is increasing on $(\mu, +\infty)$. Hence

$$\bar{\ell}(M_0 | x) = \int_0^{\mu^*} \ell(\mu^*, \mu) \pi(\mu | x) d\mu = \tau \int_0^{\mu^*} h(\mu | \mu^*) \pi(\mu | x) d\mu. \quad (2.3)$$

As Bernardo (2005a, 2010) proposes, we could agree for instance that values of $\bar{\ell}(M_0 | x)$ around $\log 10$ suggest mild evidence against H_0 , values around $\log 100$ suggest strong evidence against H_0 , and values larger than $\log 1000$ may be safely used to reject H_0 . We will have a look at the frequentist significance

level of the intrinsic hypothesis test for $\ell^* \in \{\log 10, \log 100, \log 1000\}$ in case of the reference prior, and we will compare the intrinsic hypothesis test with the Bayes factor test in case of informative priors.

Case of the reference prior. For each example we consider $M_0 = (\mu^*; \infty)$ with $\mu^* = 0.5$. We have numerically checked on numerous examples that the power function $\mu \mapsto P(\bar{\ell}(M_0 | x) > \ell^* | \mu)$ is decreasing. Hence, admitting this is indeed always the case, the test is unbiased and its significance level is the value of $P(\bar{\ell}(M_0 | x) > \ell^* | \mu)$ at $\mu = \mu^*$.

Figure 7 shows the significance level of the reference intrinsic hypothesis test in function of τ for $\ell^* = \log 10$ (mild evidence), $\ell^* = \log 100$ (strong evidence), $\ell^* = \log 1000$ (safe to reject H_0). For $\ell^* = \log 10$, the significance level is less than 3% and varies between 2% and 3% when $\tau > 50$. For $\ell^* = \log 100$, the significance level is less than 2.5‰ and varies between 1‰ and 2.5‰ when $\tau > 50$. For $\ell^* = \log 1000$, the significance level is less than 2‰ and varies between 0.8‰ and 2‰ when $\tau > 50$.

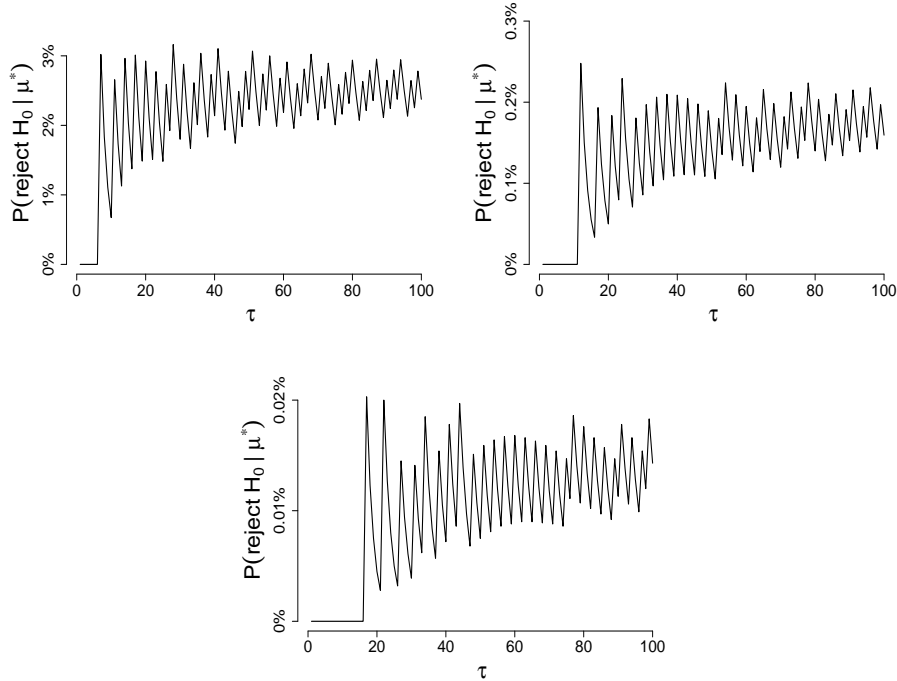


Figure 7: Significance level of the reference intrinsic hypothesis test for $H_0: \{\mu \geq 0.5\}$. *Top left:* $\ell^* = \log 10$. *Top right:* $\ell^* = \log 100$. *Bottom:* $\ell^* = \log 1000$.

Using a mathematical software, we have found that when τ is very large, the significance level approximately goes to 2.9%, 2.1‰, and 1.7‰ in case when $\ell^* = \log 10$, $\ell^* = \log 100$, $\ell^* = \log 1000$ respectively, and we have also found similar results for numerous other values of μ^* . Then we have checked that

for various values of τ , the power of the reference intrinsic hypothesis test is generally almost equal to the power of the exact Poisson test (Garwood 1936) when using these values as nominal significance levels.

Now, still considering $M_0 = (\mu^*; \infty)$ with $\mu^* = 0.5$, we compare the values of the posterior loss $\bar{\ell}(M_0 | x)$ and the values of the posterior probability $\pi(\mu \notin M_0 | x)$. For $\tau = 50, 100, 200$, figure 8 shows the results for values of x satisfying $\log 10 \leq \bar{\ell}(M_0 | x) \leq \log 1000$. We see that, approximately, the null hypothesis is rejected when $\pi(\mu \notin M_0 | x) \geq 97\%$ for $\ell^* = \log 10$, when $\pi(\mu \notin M_0 | x) \geq 99.8\%$ for $\ell^* = \log 100$, and when $\pi(\mu \notin M_0 | x) \approx 100\%$ for $\ell^* = \log 1000$. Hence the range of $\bar{\ell}(M_0 | x)$ is more convenient than the range of $\pi(\mu \notin M_0 | x)$ in order to quantify some high evidence against H_0 , because the values of $\bar{\ell}(M_0 | x)$ are more distinguishable than the values of $\pi(\mu \notin M_0 | x)$.

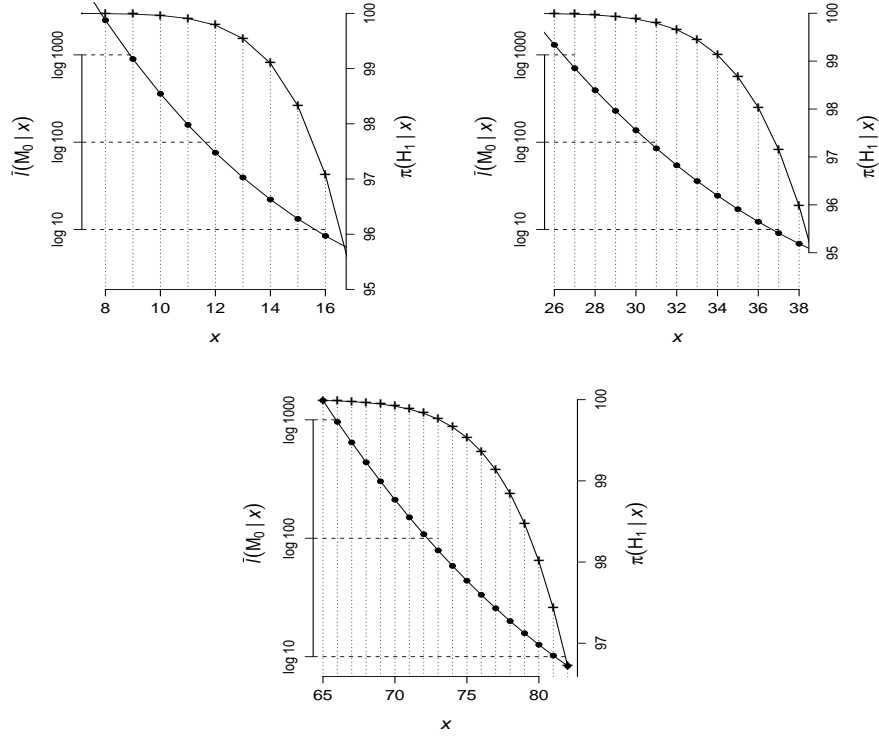


Figure 8: ‘•’(left-axis): posterior loss $\bar{\ell}(M_0 | x)$. ‘+’(right-axis): posterior probability $\pi(\mu \notin M_0 | x)$. *Top left:* $\tau = 50$. *Top right:* $\tau = 100$. *Bottom:* $\tau = 200$.

Case of an informative prior. We will firstly note that when we use an informative prior, the intrinsic hypothesis test for $H_0: \{\mu \in M_0\}$ can yield very different results than the Bayesian test based on the posterior probability of M_0 . Contrary to the intrinsic credible interval, two scenarios (τ, a, b, x) and (τ', a', b', x') yielding the same posterior distribution can yield different decisions of the intrinsic hypothesis test.

As we have just seen, when we use the reference prior, the Bayesian intrinsic hypothesis test for $H_0: \{\mu \in M_0\}$ with $M_0 = (\mu^*, \infty)$ yields similar results than the test whose rejection rule consists in rejecting H_0 if $\pi(\mu \notin M_0 | x) > p^*$ for an adequate value of p^* . This does not remain true when using an informative prior; this is particularly striking from (2.3) when we consider two scenarios with different sample sizes but with the same posterior distribution, such as the following ones already encountered in figure 5:

- $\tau = 300$, we use the reference prior ($a = 0.5, b = 0$) and we observe $x = 116$ cases;
- $\tau = 50$, we use the $\mathcal{G}(a, b)$ prior with $a = 100.5$ and $b = 250$ and we observe $x = 16$ cases.

For each case, the posterior distribution on μ is $\mathcal{G}(116.5, 300)$, and the posterior probability $\pi(\mu \notin M_0 | x)$ is approximately 99.8%. But, whereas the intrinsic credible interval is the same in both cases, we see from (2.3) that the posterior loss $\bar{\ell}(M_0 | x)$ in the first scenario equals $300/50 = 6$ times the posterior loss in the second scenario; it is approximately $\log 95$ in the first scenario, whereas it is approximately $\log 2.1$ in the second scenario. Note that if we used the reference prior in the second scenario, the posterior loss (shown in figure 8) would be approximately $\log 8.4$; since the prior predictive mean of x equals $20.1 > x^{\text{obs}} = 16$, it is sensible that the posterior loss decreases when we use this prior as compared to the noninformative prior.

Now we are going to compare the intrinsic hypothesis test with the Bayes factor test. A Bayes factor test consists in rejecting the null hypothesis $H_0: \{\mu \in M_0\}$ when the Bayes factor

$$B(x) := \left(\frac{\pi(\mu \notin M_0 | x)}{1 - \pi(\mu \notin M_0 | x)} \right) \left(\frac{\pi(\mu \notin M_0)}{1 - \pi(\mu \notin M_0)} \right)^{-1}$$

exceeds an arbitrary value b^* . We look at the posterior loss $\bar{\ell}(M_0 | x)$ and the log-Bayes factor $\log B(x)$ when we consider $\tau = 50$ and three different situations for the prior distribution:

- $\pi(\mu) = \mathcal{G}(\mu | 2k, 10k)$, whose mean is $0.2 < \mu^*$;
- $\pi(\mu) = \mathcal{G}(\mu | 5k, 10k)$, whose mean is $0.5 = \mu^*$;
- $\pi(\mu) = \mathcal{G}(\mu | 8k, 10k)$, whose mean is $0.8 > \mu^*$.

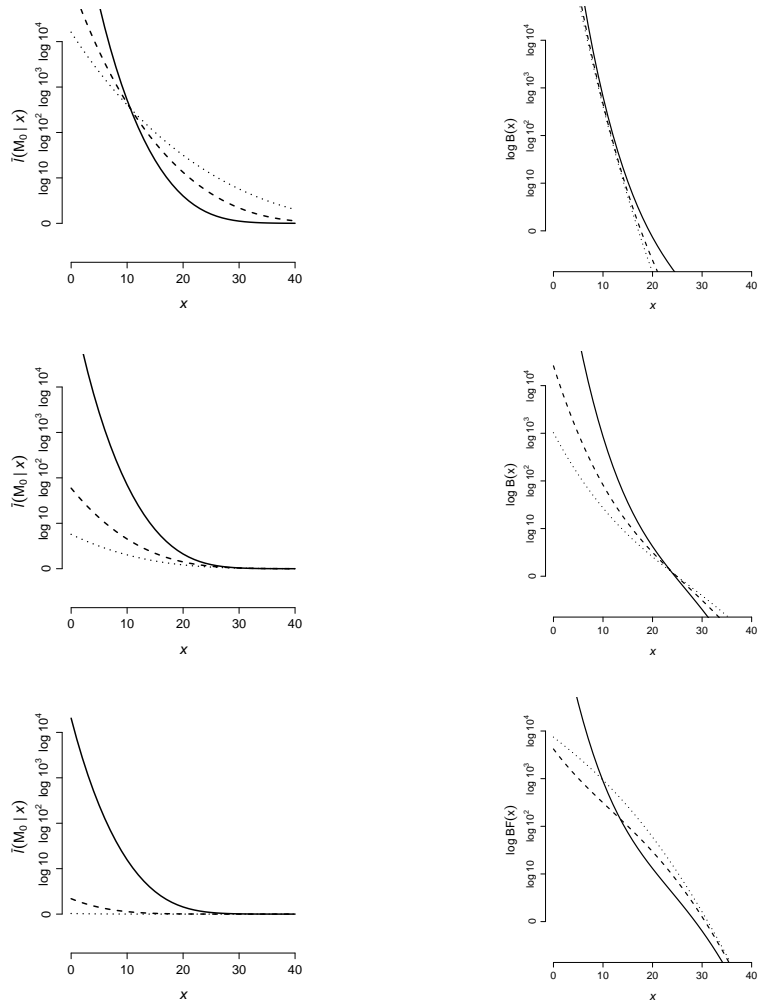


Figure 9: *Left:* Posterior loss $\bar{\ell}(M_0 | x)$. *Right:* Log-Bayes factor $\log B(x)$. *Top:* $\pi(\mu) = \mathcal{G}(\mu | 2k, 10k)$. *Middle:* $\pi(\mu) = \mathcal{G}(\mu | 5k, 10k)$. *Bottom:* $\pi(\mu) = \mathcal{G}(\mu | 8k, 10k)$. *Solid:* $k = 1$. *Dashed:* $k = 5$. *Dotted:* $k = 10$.

For each of these situations, figure 9 displays $\bar{\ell}(M_0 | x)$ and $\log B(x)$ for various values of x when $k \in \{1, 5, 10\}$. We see that the posterior loss has a very more attractive behavior than the log-Bayes factor.

3. Ratio of two Poisson rates with semi-conjugate family

In this section we consider the Bayesian model given by two independent observations $x \sim \mathcal{P}(\lambda S)$ and $y \sim \mathcal{P}(\mu T)$ where λ and μ are the unknown incidence rates and S and T are the known sample sizes, and by the following natural semi-conjugate family of priors defined by Laurent and Legrand (2011):

For any positive numbers a, b, c, d , assign the independent prior distributions on μ and $\phi := \lambda/\mu$ given by

$$\mu \sim \mathcal{G}(a, b) \quad \text{and} \quad \phi \sim \frac{T+b}{S} \times \mathcal{B}'(c, d),$$

where the Beta prime distribution $\mathcal{B}'(c, d)$ is defined by $U(1-U)^{-1} \sim \mathcal{B}'(c, d)$ when U has the Beta distribution $\mathcal{B}(c, d)$ with shape parameters c and d . Then the joint posterior on (μ, ϕ) is given by

$$(\mu \mid \phi, x, y) \sim \mathcal{G}(a+x+y, b+\phi S+T) \quad \text{and} \quad (\phi \mid x, y) \sim \frac{T+b}{S} \times \mathcal{B}'(c+x, a+d+y),$$

and it is also a proper distribution when $d = 0$ or $b = 0$.

In particular :

- when $a = c = 0.5$ and $b = d = 0$, the prior is the reference prior when ϕ is the parameter of interest (we will simply call it the reference prior, since it will always be understood that ϕ is the parameter of interest);
- when $a, b > 0$, $c = 0.5$ and $d = 0$, the prior is the semi-reference prior when the arbitrary Gamma prior distribution $\mathcal{G}(a, b)$ is assigned on μ .

The prior and posterior predictive distributions corresponding to this semi-conjugate family are provided by Laurent (2011). The “semi-reference prior” is a terminology from Laurent & Legrand; following ideas of Berger & Sun (1998), it is defined as the reference prior of the model averaged over μ with respect to its subjective prior distribution ¹.

We will derive the intrinsic credible intervals for ϕ and the intrinsic hypothesis test for null hypotheses of the form $H_0: \{\phi \in \Phi_0\}$. To do so, the *intrinsic discrepancy loss* introduced by Bernardo & Rueda (2002) is defined as follows. We denote by

$$\delta(\mu, \phi \mid \mu_0, \phi_0) := E \left(\log \frac{p(x, y \mid \mu, \phi)}{p(x, y \mid \mu_0, \phi_0)} \mid \mu, \phi \right)$$

the Kullback-Leibler divergence from the sampling distribution at (μ, ϕ) to the sampling distribution at (μ_0, ϕ_0) , and then the intrinsic discrepancy is

$$\begin{aligned} \ell(\phi_0; \phi, \mu) &:= \inf_{\mu_0 > 0} \min \{ \delta(\mu, \phi \mid \mu_0, \phi_0), \delta(\mu_0, \phi_0 \mid \mu, \phi) \} \\ &= \min \left\{ \inf_{\mu_0 > 0} \delta(\phi, \mu \mid \mu_0, \phi_0), \inf_{\mu_0 > 0} \delta(\mu_0, \phi_0 \mid \phi, \mu) \right\}. \end{aligned}$$

¹Conceptually, the model is averaged with respect to the conditional prior distribution of μ given ϕ , but here it is the same as the absolute prior distribution.

3.1. Intrinsic discrepancy loss and posterior loss

An easy direct calculation or an application of (A.2) yields

$$\delta(\mu, \phi \mid \mu_0, \phi_0) = Th \left(\mu, \frac{\phi S}{T} \mid \mu_0, \frac{\phi_0 S}{T} \right)$$

where

$$h(\mu, \gamma \mid \mu_0, \gamma_0) = \mu_0(\gamma_0 + 1) - \mu(\gamma + 1) + \mu \left(\gamma \log \frac{\gamma}{\gamma_0} + (\gamma + 1) \log \frac{\mu}{\mu_0} \right).$$

By elementary analysis, we get that $\mu_0 \mapsto h(\mu, \gamma \mid \mu_0, \gamma_0)$ has a unique minimum which is attained when $\mu(\gamma + 1) = \mu_0(\gamma_0 + 1)$ and $\mu_0 \mapsto h(\mu_0, \gamma_0 \mid \mu, \gamma)$ attains an unique minimum at $\mu_0 = \mu \exp \left(\frac{\gamma_0}{\gamma_0 + 1} \log \frac{\gamma}{\gamma_0} \right)$, and we have

$$\inf_{\mu_0 > 0} h(\mu, \gamma \mid \mu_0, \gamma_0) = \mu \left(\gamma \log \frac{\gamma}{\gamma_0} + (\gamma + 1) \log \frac{\gamma_0 + 1}{\gamma + 1} \right)$$

and

$$\inf_{\mu_0 > 0} h(\mu_0, \gamma_0 \mid \mu, \gamma) = \mu \left((\gamma + 1) - (\gamma_0 + 1) \exp \left(\frac{\gamma_0}{\gamma_0 + 1} \log \frac{\gamma}{\gamma_0} \right) \right).$$

Therefore, we finally have

$$\ell(\phi_0; \phi, \mu) = \mu T \rho \left(\frac{\phi S}{T}, \frac{\phi_0 S}{T} \right)$$

where

$$\begin{aligned} \rho(\gamma, \gamma_0) &= \min \{ f(\gamma, \gamma_0), g(\gamma, \gamma_0) \}, \\ f(\gamma, \gamma_0) &= \gamma \log \frac{\gamma}{\gamma_0} + (\gamma + 1) \log \frac{\gamma_0 + 1}{\gamma + 1} \end{aligned}$$

and

$$g(\gamma, \gamma_0) = (\gamma + 1) - (\gamma_0 + 1) \exp \left(\frac{\gamma_0}{\gamma_0 + 1} \log \frac{\gamma}{\gamma_0} \right).$$

Note that both $f(\gamma, \gamma_0)$ and $g(\gamma, \gamma_0)$ are smooth convex functions of γ_0 vanishing at $\gamma_0 = \gamma$. It is difficult to derive which of $f(\gamma, \gamma_0)$ or $g(\gamma, \gamma_0)$ equals $\rho(\gamma, \gamma_0)$ in general. Nevertheless, the following lemma, whose proof is given in Appendix B, covers some cases which may be of potential interest.

Lemma 1. 1. When $\gamma_0 = 1$, one has $\rho(\gamma, \gamma_0) = f(\gamma, \gamma_0)$ for every $\gamma > 0$.

2. When $\gamma_0 > 1$, one has $\rho(\gamma, \gamma_0) = f(\gamma, \gamma_0)$ if $\gamma \geq \gamma_0$.

3. When $\gamma_0 > e - 1$, one has $\rho(\gamma, \gamma_0) = g(\gamma, \gamma_0)$ if $0 < \gamma \leq \gamma_0$.

4. When $\gamma_0 < 1$, one has $\rho(\gamma, \gamma_0) = f(\gamma, \gamma_0)$ if $\gamma \leq \gamma_0$.

5. When $\gamma_0 < 1/(e - 1)$, one has $\rho(\gamma, \gamma_0) = g(\gamma, \gamma_0)$ if $\gamma \geq \gamma_0$.

Posterior expected intrinsic discrepancy. Given the prior of the semi-conjugate family with parameters a, b, c, d , the posterior expected intrinsic discrepancy (for short, the posterior loss) is

$$\begin{aligned}\bar{\ell}(\phi_0 | x, y) &= \iint \ell(\phi_0; \phi, \mu) \pi(\mu, \phi | x, y) d\mu d\phi \\ &= \frac{ST}{T+b} \iint \rho\left(\frac{\phi S}{T}, \frac{\phi_0 S}{T}\right) \mathcal{B}'\left(\frac{\phi S}{T+b} | c', d'\right) \mathcal{G}(\mu | a', b + \phi S + T) \mu d\mu d\phi\end{aligned}$$

where $a' = a + x + y$, $c' = c + x$ and $d' = a + d + y$. Hence,

$$\begin{aligned}\bar{\ell}(\phi_0 | x, y) &= a' \frac{ST}{T+b} \int \rho\left(\frac{\phi S}{T}, \frac{\phi_0 S}{T}\right) \mathcal{B}'\left(\frac{\phi S}{T+b} | c', d'\right) (b + \phi S + T)^{-1} d\phi \\ &= \frac{a'd'}{c' + d'} \frac{T}{T+b} \int \rho\left(\frac{\psi(T+b)}{T}, \frac{\phi_0 S}{T}\right) \mathcal{B}'(\psi | c', d' + 1) d\psi.\end{aligned}$$

In a numerical evaluation perspective, it is more convenient to express $\bar{\ell}(\phi_0 | x, y)$ as a one-dimensional integral over a finite interval:

$$\bar{\ell}(\phi_0 | x, y) = \frac{a'd'}{c' + d'} \frac{T}{T+b} \int_0^1 \rho\left(\frac{T+b}{T} \frac{u}{1-u}, \frac{\phi_0 S}{T}\right) \mathcal{B}(u | c', d' + 1) du.$$

In particular, the *reference posterior loss* is

$$\bar{\ell}(\phi_0 | x, y) = \frac{(x + y + \frac{1}{2})(y + \frac{1}{2})}{x + y + 1} \int_0^1 \rho\left(\frac{u}{1-u}, \frac{\phi_0 S}{T}\right) \mathcal{B}\left(u | x + \frac{1}{2}, y + \frac{3}{2}\right) du.$$

Comparison with the conditional model. A usual way to do statistical inference on ϕ consists in performing conditional inference given the sum $x + y$ of the two counts. The conditional distribution of x given $x + y$ is binomial with size parameter $n = x + y$ and proportion parameter

$$\theta = \frac{\lambda S}{\lambda S + \mu T} = \frac{\phi S/T}{\phi S/T + 1},$$

which is a one-to-one function of ϕ not involving the nuisance parameter μ . As noted by Laurent & Legrand (2011), this conditional binomial model has the same reference posterior on ϕ as the original unconditional model. Hence it is interesting to compare the reference posterior losses of the two models.

The reference posterior expected intrinsic discrepancy for the binomial model has been studied by Bernardo (2005b). In our context, it is the convex function $\bar{d}(\theta_0 | x, y)$ of θ_0 given by

$$\bar{d}(\theta_0 | x, y) = (x + y) \int_0^1 d(\theta_0, \theta) \mathcal{B}\left(\theta | x + \frac{1}{2}, y + \frac{1}{2}\right) d\theta,$$

where the intrinsic discrepancy loss $d(\theta_0, \theta)$ is given by

$$d(\theta_0, \theta) = \min\{\kappa(\theta | \theta_0), \kappa(\theta_0 | \theta)\}$$

and the Kullback-Leibler divergence $\kappa(\theta | \theta_0)$ is given by

$$\kappa(\theta | \theta_0) = \theta \log \frac{\theta}{\theta_0} + (1 - \theta) \log \frac{1 - \theta}{1 - \theta_0}$$

(this can be derived from (A.2)).

Figure 10 shows the reference posterior loss $\bar{\ell}(\phi_0 | x, y)$ for the original unconditional model and the reference posterior loss $\bar{d}\left(\frac{\phi_0 S/T}{1 + \phi_0 S/T} | x, y\right)$ for the conditional model, when $x = 6$, $y = 15$, and $S/T = 1$. We see on this example that the posterior losses are very close, and we have checked that this is always the case on numerous examples.

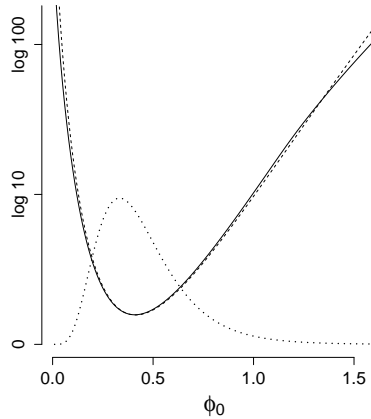


Figure 10: Reference posterior distribution on ϕ (dotted), reference posterior loss for the unconditional model (solid) and reference posterior loss for the conditional model (dashed) in function of ϕ_0 , when $x = 6$, $y = 15$, and $S/T = 1$.

3.2. Intrinsic credible intervals

Contrary to the case of section 2, the posterior loss $\bar{\ell}(\phi_0 | x, y)$ is not a convex function of ϕ_0 . We have numerically checked on numerous examples that the curve $\phi_0 \mapsto \bar{\ell}(\phi_0 | x, y)$ always has the same shape as in the example of figure 10. Admitting that this is indeed always the case, the set

$$I_{x,y} = \{\phi_0 | \bar{\ell}(\phi_0 | x, y) < c\}$$

consisting of those values of ϕ_0 for which the posterior loss does not exceed a given value c , is always an interval whatever the value of c is. When $c = c(x, y)$ is chosen in order that $\pi(\phi \in I_{x,y} | x, y) = q$ for a given $q \in (0, 1)$, then $I_{x,y}$ is the *intrinsic* $100q\%$ -credible interval of ϕ .

In the case of the reference prior, we have found on numerous examples that, similarly to the case studied in section 2, the reference intrinsic credible interval is close to the equal-tailed and HPD credible intervals, and its frequentist coverage is comparable with the one of the equal-tailed and HPD credible intervals, which is studied by Laurent & Legrand (2011). Things are not so simple in the case of an informative prior, but in most cases the intrinsic credible interval remains relatively close to the equal-tailed and HPD credible intervals.

3.3. Intrinsic hypothesis testing for one-sided hypotheses

Let $\Phi_0 \subset (0, +\infty)$. The *intrinsic hypothesis test* for $H_0: \{\phi \in \Phi_0\}$ is derived with the help of the intrinsic discrepancy $\ell(\Phi_0; \phi, \mu)$ from Φ_0 to (ϕ, μ) defined by $\ell(\Phi_0; \phi, \mu) = \inf_{\phi_0 \in \Phi_0} \ell(\phi_0; \phi, \mu)$. The intrinsic hypothesis test (called the *Bayesian reference criterion* in case of the reference prior) consists in rejecting H_0 if $\bar{\ell}(\Phi_0 | x, y) > \ell^*$ where ℓ^* is a given utility constant and

$$\bar{\ell}(\Phi_0 | x, y) = \iint \ell(\Phi_0; \phi, \mu) \pi(\mu, \phi | x, y) d\mu d\phi$$

is the posterior expected value of the intrinsic discrepancy from Φ_0 to (ϕ, μ) , briefly termed as the posterior loss.

We consider the case when $\Phi_0 = (\phi^*, +\infty)$ for a given value of ϕ^* . As the function $\phi_0 \mapsto \ell(\phi_0; \phi, \mu)$ is decreasing on $(0, \phi)$ and increasing on $(\phi, +\infty)$, then, using the prior of the semi-conjugate family with parameters a, b, c, d , the posterior loss is

$$\begin{aligned} \bar{\ell}(\Phi_0 | x, y) &= \int_0^\infty \int_0^{\phi^*} \ell(\phi^*; \phi, \mu) \pi(\mu, \phi | x, y) d\phi d\mu \\ &= \frac{a'd'}{c'+d'} \frac{T}{T+b} \int_0^{u^*} \rho\left(\frac{T+b}{T} \frac{u}{1-u}, \frac{\phi^* S}{T}\right) \mathcal{B}(u | c', d' + 1) du \end{aligned}$$

where $a' = a + x + y$, $c' = c + x$, $d' = a + d + y$ and $u^* = \frac{\phi^* S / (T+b)}{1 + \phi^* S / (T+b)}$. In the particular case of the reference prior, it depends on S and T only through S/T . Note that in the case when $\frac{\phi^* S}{T} \leq 1$, thanks to lemma 1, one has

$$\bar{\ell}(\Phi_0 | x, y) = \frac{a'd'}{c'+d'} \frac{T}{T+b} \int_0^{u^*} g\left(\frac{T+b}{T} \frac{u}{1-u}, \frac{\phi^* S}{T}\right) \mathcal{B}(u | c', d' + 1) du,$$

where g is the function defined just above lemma 1.

We now have a look at the frequentist properties of the reference intrinsic hypothesis test for $H_0: \{\phi > \phi^*\}$ when $S/T = 1$. In this case, the probability $P(\bar{\ell}(\Phi_0 | x, y) > \ell^* | \mu, \phi)$ for rejecting H_0 depends on μ and T only through $\mu T = E(y | \mu, \phi)$. We have numerically checked on numerous examples that the power function $\phi \mapsto P(\bar{\ell}(\Phi_0 | x, y) > \ell^* | \mu, \phi)$ is decreasing, hence the test is unbiased for a fixed value of μ and we can consider that the significance level is the value of $P(\bar{\ell}(\Phi_0 | x, y) > \ell^* | \mu, \phi)$ at $\phi = \phi^*$. The significance level is displayed on table 1 and figure 11. We see that for a fixed value of ϕ^* ,

the significance level is practically increasing with μT and stabilizes around a certain value, which is approximately 2.8% when $\ell^* = \log 10$ (mild evidence), 2⁰/₀₀ when $\ell^* = \log 100$ (strong evidence), and 1.6⁰/₀₀₀ when $\ell^* = \log 1000$ (safe to reject H_0).

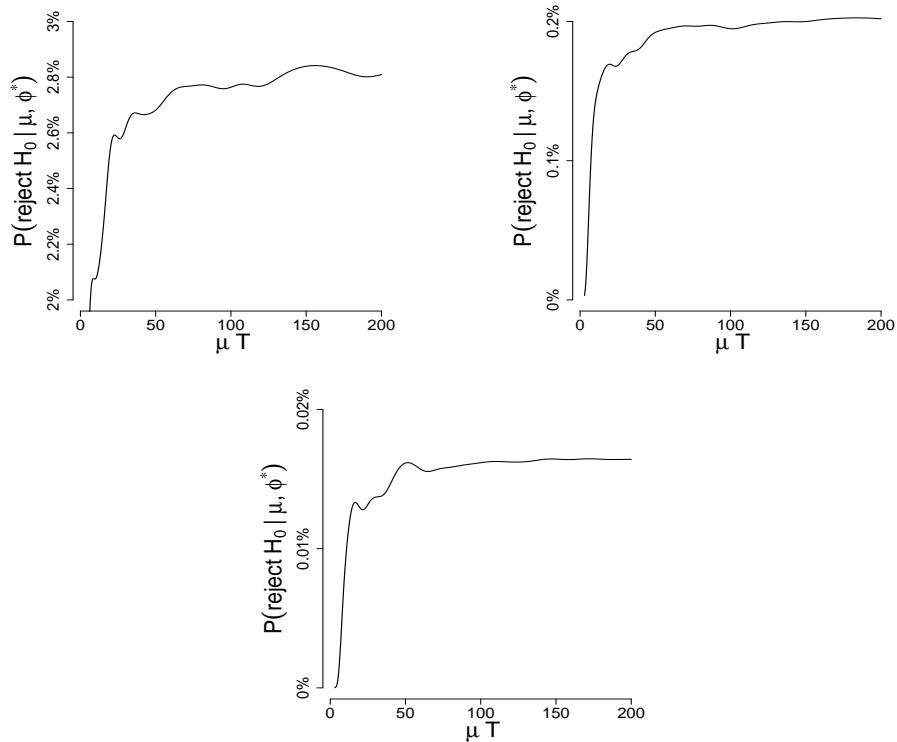


Figure 11: Significance level of the reference intrinsic test for $H_0: \{\phi > 0.75\}$ in function of μT when $S = T$. *Top left:* $\ell^* = \log 10$. *Top right:* $\ell^* = \log 100$. *Bottom:* $\ell^* = \log 1000$.

We do not display some results on the power achieved by the reference intrinsic test, since we have checked on numerous examples that, analogously to the case of the reference intrinsic test for one Poisson rate, it is approximately the same as the power achieved by the test whose rejection rule consists in rejecting H_0 if $\pi(\phi \notin \Phi_0 | x, y) > p^*$ when p^* is chosen in order that the significance levels are approximately the same, and the power of this test is studied by Laurent & Legrand (2011).

4. Conclusion

We think that intrinsic Bayesian inference, owing to its sensible and coherent foundations and behavior, has a great value and enjoys a potential popularization; this one may require that practitioners familiarize themselves with the

(a) $\ell^* = \log 10$, values given in %								
ϕ^* \backslash μT	2	4	10	15	20	30	40	50
0.1	0.0	0.0	0.0	0.0	0.0	0.8	1.5	1.7
0.3	0.0	0.0	1.1	1.6	1.9	2.1	2.4	2.4
0.5	0.0	0.3	2.0	2.3	2.5	2.5	2.5	2.6
0.75	0.4	1.4	2.1	2.2	2.5	2.6	2.7	2.7
1	0.8	1.9	2.6	2.6	2.7	2.8	2.7	2.8

(b) $\ell^* = \log 100$, values given in ‰								
ϕ^* \backslash μT	2	4	10	15	20	30	40	50
0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3
0.3	0.0	0.0	0.0	0.4	0.9	1.2	1.4	1.5
0.5	0.0	0.0	0.7	1.2	1.4	1.5	1.7	1.8
0.75	0.0	0.2	1.4	1.6	1.7	1.8	1.8	1.9
1	0.0	0.6	1.4	1.7	1.8	1.9	2.0	2.0

(c) $\ell^* = \log 1000$, values given in ‰ ₀₀₀								
ϕ^* \backslash μT	2	4	10	15	20	30	40	50
0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.3	0.0	0.0	0.0	0.0	0.2	0.9	1.0	1.1
0.5	0.0	0.0	0.1	0.6	0.9	1.1	1.3	1.4
0.75	0.0	0.0	0.9	1.3	1.3	1.4	1.5	1.6
1	0.0	0.2	1.2	1.2	1.4	1.5	1.6	1.6

Table 1: Significance level of the reference intrinsic test for $H_0: \{\phi > \phi^*\}$ when $S/T = 1$.

concept of bit of information as a unit of divergence between two models. In Bayesian statistics, the posterior distribution authorizes a quantitative discourse on the unknown parameter, and this one is considerably enriched by the posterior expected intrinsic discrepancy loss. The main points emerging from our work are the following.

The intrinsic credible interval is close to the HPD and equal-tailed credible intervals, but it has the advantage to be associated to a lowest loss level: it gives not only a credible region but also a measure of compatibility of this region with the observed data.

In case when we use the reference prior, the intrinsic credible interval as well as the intrinsic hypothesis test enjoy attractive frequentist properties. The reference intrinsic hypothesis test yields decisions close to the test based on the posterior probability of the null hypothesis but the range of the posterior loss is more convenient to calibrate the degree of evidence against the null hypothesis.

In case when we use an informative prior, the intrinsic hypothesis test behaves more sensibly than the test based on the posterior probability of the null hypothesis; we have seen that the intrinsic hypothesis test takes also into

account the sample size in addition to the posterior distribution. These two tests can yield different conclusions, and this might call into question the standard practice of subjective Bayesian statistics. Intrinsic hypothesis testing also behaves more sensibly than Bayes factors.

Concerning the computations, we have seen that they are more difficult with the intrinsic discrepancy loss than with the entropy loss (Robert 1996); we refer to Bernardo (2005a, 2010) for theoretical arguments in favor of the intrinsic discrepancy. These computational difficulties motivated Bernardo (2005a, 2010) to provide approximations of the reference posterior loss for large sample sizes.

Appendix A. Intrinsic loss in the exponential family

Let us consider a parametric family $\{p_\theta\}$ of distributions on \mathbb{R}^k with parameter $\theta \in \mathbb{R}^k$ such that p_θ has a density of the form

$$f(x \mid \theta) = e^{\theta \cdot x - \psi(\theta)} \quad (\text{A.1})$$

with respect to some dominating measure ν . For instance, in the context of sections 2 and 3:

- the Poisson distribution $\mathcal{P}(\mu\tau)$ has a density of the form (A.1) with $\theta = \log(\mu\tau)$, $\psi(\theta) = \mu\tau = e^\theta$, and ν is the measure assigning mass $1/x!$ at point $x \in \mathbb{N}$.
- the independent product $\mathcal{P}(\lambda S) \otimes \mathcal{P}(\mu T)$ has a density of the form (A.1) with $\theta = (\log(\lambda S), \log(\mu T))$, $\psi(\theta) = e^{\theta_1} + e^{\theta_2}$, and ν is the measure assigning mass $1/(x!y!)$ at point $(x, y) \in \mathbb{N}^2$.

In case when ψ is differentiable, Robert (1996) showed that the Kullback-Leibler divergence from p_θ to p_{θ_0}

$$\delta(\theta \mid \theta_0) := E \left(\log \frac{f(x \mid \theta)}{f(x \mid \theta_0)} \mid \theta \right)$$

is given by

$$\delta(\theta \mid \theta_0) = (\theta - \theta_0) \cdot \nabla \psi(\theta) + \psi(\theta_0) - \psi(\theta). \quad (\text{A.2})$$

Appendix B. Proof of lemma 1

With the notations of lemma 1, we put $F(x, y) = f(x, y) - g(x, y)$ for $x > 0$ and $y > 0$. One has

$$\frac{\partial^2 F}{\partial x^2}(x, y) = \frac{N(x, y)}{D(x, y)}$$

where $D(x, y) > 0$ and

$$N(x, y) = x(1+y) - y(1+x) \left(\frac{x}{y} \right)^{\frac{y}{1+y}}.$$

By elementary analysis, we show that, when x varies from 0 to ∞ , the function $x \mapsto N(x, y)$ is firstly decreasing, then increasing, and finally decreasing. Moreover, one has $N(0, y) = N(y, y) = 0$, and it is easy to see that $N(x, y) \leq 0$ when $x < y \leq 1$ or $x > y \geq 1$. Then we deduce that $F^{(10)}(x, y) := \frac{\partial F}{\partial x}(x, y)$ which is given by

$$F^{(10)}(x, y) = -1 + \left(\frac{x}{y}\right)^{-\frac{1}{1+y}} + \log\left(\frac{x}{y}\right) + \log\left(\frac{1+y}{1+x}\right),$$

and which satisfies $F^{(10)}(0, y) = +\infty$, $F^{(10)}(y, y) = +\infty$ and $F^{(10)}(+\infty, y) = \log\left(1 + \frac{1}{y}\right) - 1$, has the following properties:

- when $y = 1$, the function $x \mapsto F^{(10)}(x, y)$ decreases and vanishes at $x = y$;
- when $y > 1$, the function $x \mapsto F^{(10)}(x, y)$ is decreases from 0 to some value $a(y) < y$ where it is negative, and it remains negative for $x \geq a(y)$;
- when $y < 1$, the function $x \mapsto F^{(10)}(x, y)$ is positive from 0 to some value $b(y) > y$, and it is decreasing on $(b(y), +\infty)$, hence in particular $F^{(10)}(x, y) \geq 0$ for every $x \in (0, +\infty)$ when $y \leq 1/(e-1)$.

Lemma 1 follows from the facts that $F(0, y) = \log(1+y) - 1$ and $F(+\infty, y) = -\infty$ or $F(+\infty, y) = +\infty$ according to $y > 1/(e-1)$ or $y \leq 1/(e-1)$ respectively.

Acknowledgements

This research was partially supported by IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy) while the author was a postdoctoral fellow at Université catholique de Louvain.

References

References

- [1] Berger JO and Sun D. 1998. Reference priors with partial information. *Biometrika* 85(1):55–71.
- [2] Bernardo JM. 1999. Nested hypothesis testing: The Bayesian reference criterion (with discussion). In *Bayesian Statistics 6*, eds. Bernardo JM, Berger JO, Dawid AP and Smith AFM, 101–130. Oxford: Oxford University Press.
- [3] Bernardo JM. 2005a. Reference Analysis. In *Handbook of Statistics 25*, eds. Dey DK and Rao CR, 17–90. Amsterdam: Elsevier.
- [4] Bernardo JM. 2005b. Intrinsic Credible Regions: an Objective Bayesian approach to Interval Estimation (with discussion). *Test* 14:317–384.
- [5] Bernardo JM. 2009. Modern Bayesian Inference: Foundations and Objective Methods. In *Philosophy of Statistics*, eds. Bandyopadhyay P and Forster M. Amsterdam: Elsevier.
- [6] Bernardo JM. 2010. Integrated objective Bayesian Estimation and hypothesis testing (to appear, with discussion). In *Bayesian Statistics 9*, eds. Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AFM and West M. Oxford: Oxford University Press.
- [7] Bernardo JM and Juarez MA. 2003. Intrinsic estimation. In *Bayesian Statistics 7*, eds. Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AFM and West M, 465–476. Oxford: Oxford University Press.
- [8] Bernardo JM and Rueda R. 2002. Bayesian hypothesis testing: A reference approach. *Internat. Statist. Rev.* 70:351–372.
- [9] Bernardo JM and Smith AFM. 1994. *Bayesian Theory*. Chichester: Wiley.
- [10] Datta GS and Sweeting TJ. 2005. Probability matching priors. In *Handbook of Statistics 25*, eds. Dey DK and Rao CR, 91–114. Amsterdam: Elsevier.
- [11] Garwood F. 1936. Fiducial Limits for the Poisson Distribution. *Biometrika* 28:437–442.
- [12] Laurent S. 2011. *Some Poisson mixtures distributions with a hyperscale parameter*. To appear in: *Brazilian Journal of Probability and Statistics*.

- [13] Laurent S and Legrand C. 2011. A Bayesian framework for the ratio of two Poisson rates in the context of vaccine efficacy trials. To appear in: *ESAIM: P&S*.
- [14] Robert C. 1996. Intrinsic losses. *Theory and Decision* 40:191–214.
- [15] Robert C. 2007. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. 2nd ed. New-York: Springer-Verlag, Springer Texts in Statistics.