Bayesian Analysis (2014) **9**, Number 4, pp. 939–962

Equivalence between the Posterior Distribution of the Likelihood Ratio and a p-value in an Invariant Frame.

Isabelle Smith *† and André Ferrari ‡

Abstract. The Posterior distribution of the Likelihood Ratio (PLR) is proposed by Dempster in 1973 for significance testing in the simple vs. composite hypothesis case. In this hypothesis test case, classical frequentist and Bayesian hypothesis tests are irreconcilable, as emphasized by Lindley's paradox, Berger & Selke in 1987 and many others. However, Dempster shows that the PLR (with inner threshold 1) is equal to the frequentist p-value in the simple Gaussian case. In 1997, Aitkin extends this result by adding a nuisance parameter and showing its asymptotic validity under more general distributions. Here we extend the reconciliation between the PLR and a frequentist p-value for a finite sample, through a framework analogous to the Stein's theorem frame in which a credible (Bayesian) domain is equal to a confidence (frequentist) domain.

Keywords: hypothesis testing, PLR, p-value, likelihood ratio, frequentist and Bayesian reconciliation, Lindley's paradox, invariance

1 Introduction

1.1 Classical hypothesis test methodologies

Simple versus composite hypothesis testing is a general statistical issue in parametric modeling. It consists for a given observed dataset x in choosing among the hypothesis

$$\mathrm{H}_0 : \theta = \theta_0 \qquad \mathrm{H}_1 : \theta \in \Theta_1 \tag{1}$$

where the distribution of x is characterized by the underlying unknown parameter θ. Under the alternative hypothesis H_1, θ takes a value different from the point θ_0, and the uncertainty of θ is described by a prior probability density function $\pi_1(\theta)$ which is positive only for $\theta \in \Theta_1$. We assume that the data model $p(x|\theta)$ has the same expression under H_0 and H_1.

To choose among H_0 and H_1, a test statistic $T(x)$ (such as the Generalized Likelihood Ratio) is generally compared to a threshold ζ and one decides to choose H_0 if $\mathrm{T}(x)$ is greater than ζ. If H_1 is chosen whereas the true underlying θ was equal to θ_0, a type I

*Laboratoire des Sciences du Climat et de l'Environnement ; IPSL-CNRS, Orme des Merisiers, 91191 Gif-sur-Yvette, France.
†Laboratoire H. Fizeau ; Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur, France. zazoo@mac.com
‡Laboratoire J.-L. Lagrange ; Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur, France. andre.ferrari@unice.fr

 DOI:10.1214/14-BA877

error is made in the decision. Under the classical Neyman paradigm (see Neyman and Pearson (1933); Neyman (1977)), the threshold ζ is chosen so that the probability of the type I error lies under (or is equal to) some fixed level α, typically a 5% error rate. Instead of inverting this function, a p-value can be defined in order to serve as the test statistic to be directly compared to the 5% level (Lehmann and Romano (2005)):

$$p_{\mathrm{val}}(\mathrm{T}(x_0)) = \Pr(\mathrm{T}(x_0) < \mathrm{T}(x)|\theta_0) \tag{2}$$

where x_0 is the observed dataset and x the variable of integration. Note that with this notation, H_0 is rejected when $p_{\mathrm{val}}(\mathrm{T}(x_0))$ is greater than some threshold.

On the Bayesian side, the test statistic classically used (Robert (2007)) is the Bayes Factor (BF) defined by

$$\mathrm{BF}(x) = \frac{p(x|\theta_0)}{\int d\theta \; p(x|\theta)\pi_1(\theta)} \; .$$

Making a binary decision consists of choosing H_0 if $\mathrm{BF}(x)$ is greater than some threshold, and the choice of the threshold is made in general by a straight interpretation of the BF. The Jeffreys's scale for example states that if the observed BF is between 10 and 100 there is a strong evidence in favor of H_0. The mere posterior probability $\Pr(\mathrm{H}_i|x)$ of an hypothesis may also be considered by itself.

A practical issue of the BF in the simple vs. composite hypothesis test is that it is defined up to a multiplicative constant if the prior π_1 is improper[1] even though the posterior distribution is proper. Partial BFs account for this issue by somehow using part of the data to update the prior into a proper posterior, and then using this posterior as the prior for the rest of the data. The most simply defined Partial BF is the Fractional BF (FBF) proposed by O'Hagan (1995).

A related and more fundamental issue is Lindley's paradox, initially studied by Jeffreys (1961) and called a paradox by Lindley (1957), which shows among others that, when testing a simple vs. a composite hypothesis, the null hypothesis H_0 is too highly favoured against H_1 for a natural diffuse prior under Θ_1. More precisely, for example in the test of the mean of a Gaussian likelihood, the p-value $|x|$ defines the uniformly most powerful test, which is a very strong optimal property even according to at least part of the Bayesian community. However, for a fixed prior and some dataset x that adjusts so that the associated classical p-value remains fixed (so that the evidence for H_0 shall not change), $\Pr(\mathrm{H}_0|x)/\Pr(\mathrm{H}_1|x)$ tends to 1 as the sample size increases. This issue, intensively discussed and developed (see Tsao (2006) for a quite recent study), is consensually considered as a real trouble by a quite large part of the community. Unlike the BF, other tests like the FBF or the Bernardo (2011) test do not suffer from this problem in Lindley's frame. Other ideas have been developed which prevent Lindley's frame from occurring, avoiding troubles for the BF. Berger and Delampady (1987) for example argue that testing a simple hypothesis is an unreasonable question. Some other references will be given in the section 2.1.

[1] π_1 is called improper if its integral over Θ_1 is infinite, which occurs if π_1 is constant over an unbounded domain for example.

Among many frequentist and Bayesian p-values (several are listed by Robins et al. (2000)), the next most classical Bayesian-type hypothesis test statistic is the posterior predictive p-value, highlighted by Meng (1994). Unlike the BF which only integrates over the parameter space Θ, the posterior predictive p-value integrates over the data space $\mathcal{X}$, like frequentist p-values. But unlike the frequentist p-value which integrates under the frequentist likelihood $p(x|\theta_0)$, it integrates under the predictive likelihood $p(x^{\text{pred}}|x_0) = \int d\theta\, p(x^{\text{pred}}|\theta)\pi(\theta|x_0)$ where x_0 is the observed dataset. In a frequentist p-value only a statistic (i.e. a function of x only) can define the domain of integration. On the contrary, in the posterior predictive p-value, a discrepancy variable (function of both x and θ) can be used to define the domain of integration. Note that the choice of the discrepancy variable to use there remains an issue.

Although a bit less classical, the approach of Evans (1997) needs to be introduced because the tool and some of its properties are interesting and closely related to the ones derived in this paper. In the simple vs. composite test case presented up to now, the tool proposed by Evans (1997) and the ones studied in this paper are even mathematically equal. But the tool proposed by Evans (1997) is defined to test more generally $\mathrm{H}_0 : \Psi(\theta) = \psi_0$ for a parameter of interest $\psi = \Psi(\theta)$. The test statistic consists of measuring the Observed Relative Surprise (ORS) related to the hypothesis by computing:

$$\mathrm{ORS}(x) = \Pr\left(\frac{\pi_\Psi(\Psi(\theta)|x)}{\pi_\Psi(\Psi(\theta))} \geq \frac{\pi_\Psi(\psi_0|x)}{\pi_\Psi(\psi_0)} \middle| x \right) . \tag{3}$$

The relative belief ratio of ψ defined by $\mathrm{RB}(\psi) = \pi_\Psi(\Psi(\theta)|x)\pi_\Psi(\Psi(\theta))^{-1}$ is measuring the change in belief in ψ being the true value of $\Psi(\theta)$ from *a priori* to *a posteriori*. So if $\mathrm{RB}(\psi_0) > 1$ we have evidence in favor of H_0. Relative belief ratios are discussed in Baskurt and Evans (2013) where $\mathrm{RB}(\psi_0)$ is presented as the evidence for or against H_0 and (3) is presented as a measure of the reliability of this evidence. This leads to a possible resolution of Lindley's paradox as the relative belief ratio can be large and ORS small without contradiction. See Example 4 of Baskurt and Evans (2013) and note that Evans (1997) shows that ORS converges to the classical p-value as the prior becomes more diffuse in this example.

1.2 Posterior distribution of the Likelihood Ratio (PLR)

Let's focus again on the simple vs. composite hypothesis test. Contrary to the posterior predictive p-value, the Posterior distribution of the Likelihood Ratio (PLR) does not integrate over some data which are unobserved, but only integrates over Θ. It still conditions upon the only observed variable, namely x_0, like for the BF, but on a domain defined from a divergence variable, like the posterior predictive p-value. This statistic proposed by Dempster (1973) is defined by

$$\mathrm{PLR}(x,\zeta) = \Pr\big(\mathrm{LR}(x,\theta) \leq \zeta \big| x\big) \tag{4}$$

where $\mathrm{LR}(x, .)$ is the Likelihood Ratio

$$\mathrm{LR}(x, \theta) = \frac{p(x|\theta_0)}{p(x|\theta)} \quad \theta \in \Theta_1 \ .$$

Since θ is random, the deterministic function $\mathrm{LR}(x, .)$ evaluated at the random variable θ becomes naturally random with some posterior distribution characterized by its cumulative distribution, the PLR. As emphasized by Birnbaum (1962), Dempster (1973) and Royall (1997), the threshold ζ which compares the original likelihoods under H_0 and under H_1 is directly interpretable and can be chosen the same way an error level α is chosen in the Neyman-Pearson paradigm. "$\mathrm{PLR}(x, 1) = 0.1$" for example reads "The probability that the likelihood of θ_1 is more than the likelihood of θ_0 is 0.1.".

The PLR can therefore be used for a binary decision, by fixing ζ and deciding to reject H_0 if $\mathrm{PLR}(x, \zeta)$ is greater than, say, 0.9. One can check if the binary decision is sensitive to the choice of both thresholds by making the test for several thresholds and see if the decision is different. In the extreme case, note that due to the nice definition of the PLR, one can simply display $\mathrm{PLR}(x, \zeta)$ as a function of ζ to get a broad view. The range of ζ under which $\mathrm{PLR}(x, \zeta)$ grows, typically from 0.2 to 0.8, indicates if the decision for H_0 or H_1 is clear, or not. As soon as the posterior can be sampled, these computations and graphs are very easy to display as will be explained later.

The PLR has been first proposed by Dempster (1973, 1997), then studied especially by Aitkin (1997) and Aitkin (2010) but also used and analyzed by Aitkin et al. (2005, 2009). As mentioned in the previous subsection, it turns out that the PLR is also closely related to the ORS proposed by Evans (1997), which generalizes the PLR. The PLR is also closely related to the e-value associated to the Full Bayesian Significance Test (FBST) from Pereira and Stern (1999) and slightly revisited by Borges and Stern (2007) which then somehow generalizes the PLR by adding a reference distribution on θ, and by systematically dealing with the case where the null hypothesis domain Θ_0 has a dimension less than Θ_1 but which is not necessarily restricted to the point $\Theta_0 = \{\theta_0\}$. We do not list the results found by these different analyses, apart from some specifically mentioned ones.

The PLR turns out to be a natural Bayesian measure of evidence of the studied hypothesis since it involves only the posterior distribution of θ (no integral over $\mathcal{X}$) and the likelihood, claimed by Birnbaum (1962), Royall (1997) and others to be the only tool that can measure evidence. Unlike the BF, the PLR is well defined for an improper prior as soon as the posterior is proper, and is not subject to Lindley's paradox. It is also invariant under any isomorphic transformation of the $\mathcal{X}$ space and any transformation of the Θ space, as a consequence of being a mere function of the likelihood. These last properties were emphasized for example for the e-value associated to the FBST.

The PLR also consists of a natural alternative to the BF in different regards. To start with, the PLR first compares (compares $p(x|\theta_0)$ and $p(x|\theta)$) and then integrates, whereas the BF first integrates and then compares (compares $p(x|\theta_0)$ and $\int d\theta \ p(x|\theta)\pi(\theta)$). Second, Newton and Raftery (1994) and many others show that if the prior under H_1 is proper, the BF is simply the posterior mean of the LR, i.e. the mean of the distribution described by the PLR[1]. However a point estimate is in general not given alone but accompanied by an uncertainty indicator. Smith and Ferrari (2010) show that the posterior mean of the LR raised to some power is equal to the FBF introduced previously; the mean of the PLR is given by the BF and its variance is easily related to the FBF. However, Smith (2010) shows that the Generalized Likelihood Ratio bounds the support (values of $\zeta(x)$ for which $\mathrm{PLR}(x,\zeta) > 0$) of the PLR and that at this lower bound the PLR in general starts by an infinite derivative. In addition to this theoretical result, numerical examples also indicate that the posterior density function of the LR is in general highly asymmetric. Therefore, the BF (point estimate of the LR) or any standard centered credible intervals do not appear to be relevant inferences about the LR seen as a random variable. Instead, the same way the BF is to be thresholded, the actual information about $\mathrm{LR}(x,\theta)$ which seems to be relevant, and invariant under the transformation $\mathrm{LR}(x,\theta) \mapsto (\mathrm{LR}(x,\theta))^{-1}$, is to indicate its cumulative posterior distribution, which is precisely the PLR.

In practice, the PLR can be straightforwardly computed as soon as the posterior distribution $\pi(\theta|x)$ is sampled. Just obtain from a Monte Carlo Markov Chain (MCMC) algorithm an almost i.i.d. chain $\{\theta^{[1]}, ..., \theta^{[m]}\}$ from the posterior distribution $\pi(\theta|x)$ and compute $\mathrm{LR}(x, \theta^{[i]})$ for each sample. The resulting histogram sketches the posterior density of the LR and the plot of the empirical cumulative distribution of the LR chain sketches the PLR as a function of ζ.

The PLR has been realistically and thoroughly applied by Smith (2010) to the detection of extra-solar planets from images acquired with the dedicated instrument SPHERE mounted on the Very Large Telescope. At this moment, only very finely simulated images were available. The PLR has been applied to two simulated datasets, one in which no extra-solar planet is present (dataset simulated under H_0) and the other in which an extra-solar planet is present (H_1 dataset). Although the extra-solar planet is very dark (10^6 times less bright than the star it surrounds), close to the star (angular distance in the sky of 0.2 arcseconds i.e. 6.10^{-5} degrees), and although only 2×20 images were used, thanks to the quality of the optical instruments and of the statistical model the detection and not detection were evident, with $\mathrm{PLR}(x, 0.1) = 0.0$ for the dataset under H_0 and $\mathrm{PLR}(x, 0.1) = 0.94$ for the dataset under H_1. As studied by Smith (2010), the statistical model and consecutive method are very satisfying compared to classical methods.

1.3 Problems addressed here

Despite its potential interest the PLR has not been extensively studied up to now. This paper aims at contributing in this investigating work by some new results.

In the simple vs. composite hypothesis test case, it turns out that the PLR plays a strong role in understanding the possible reconciliation between frequentist and Bayesian hypothesis testing. The PLR with inner threshold $\zeta = 1$ is simply equal to some

[1]Alternatively, note that if we had defined the BF and LR with the alternative hypothesis at the numerator of these fractions, the BF would have been the prior mean of the LR.

frequentist p-value for some "likelihood - prior - hypothesis" combinations. Dempster (1973) and Aitkin (1997) first noticed and highlighted this equivalence when testing the mean of a Gaussian likelihood with a uniform prior.

In Section 2, we extend the conditions of this equivalence result under a frame analogous to the one used to reconcile confidence and credible domains. Subsection 2.1 synthesizes the long quest of reconciliation between frequentist and Bayesian hypothesis tests, Subsection 2.2 proves and discusses the reconciliation reached between the PLR and some frequentist p-value in such an invariant frame, Subsection 2.3 gives examples and perspectives, and Subsection 2.4 discusses the connection between this reconciliation result and the one obtained between (frequentist) confidence domains and (Bayesian) credible regions.

A concluding discussion is proposed in Section 3. The appendices essentially present the proofs of the mathematical results.

2 Equivalence between the PLR and a frequentist p-value

2.1 Previous tentative reconciliations of frequentist and Bayesian tests

As introduced in Section 1.1, Lindley's paradox presents a frame where $\Pr(\mathrm{H}_0|x)$ (often thought as being *the* Bayesian measure of evidence) may be expected to be equal to the frequentist p-value, but happens not to be. Also, the BF is not satisfying in the frame "point null hypothesis H_0 and diffuse prior π_1". This highlights the need for other Bayesian-type hypothesis tests, but also raises more generally the question of reconciliation between frequentist and Bayesian hypothesis tests.

The conditions upon which frequentist (Neyman (1977)) and Bayesian (Jeffreys (1961)) answers agree are always of interest in order to understand the interpretation of the procedures and the limits of the two paradigms, somehow defined by what they are *not*.

A first approach to see when could frequentist and Bayesian hypothesis tests be unified consists of analyzing, for different hypothesis likelihoods and priors, when are the classical p-value and $\Pr(\mathrm{H}_0|x)$ equal. These two concepts are to be compared because they both seem to handle only H_0 and in very simple ways, one from the frequentist the other from the Bayesian perspectives[1]. It turns out that unlike for a composite null hypothesis (e.g. Casella and Berger (1987)), for a point null hypothesis Lindley's paradox $\Pr(\mathrm{H}_0|x) > p_{\mathrm{val}}(x)$ always seems to hold. Berger and Sellke (1987) in particular show that among very broad classes of priors $\Pr(\mathrm{H}_0|x) > p_{\mathrm{val}}(x)$ always holds for $\Pr(\mathrm{H}_0) = 0.5$. Also see the extensive list of references included. Oh and DasGupta (1999) follows this analysis by studying the effect of the choice of $\Pr(\mathrm{H}_0)$.

Another approach consists of modifying the standard frequentist procedure and/or

[1]Note however that H_1 is implicitly taken into account through the marginal distribution of x in $\Pr(\mathrm{H}_0|x)$.

the standard Bayesian hypothesis test procedure, but still relying on the p-value and $\Pr(\mathrm{H}_0|x)$, to see if they can then be made equivalent. Berger and Delampady (1987) for example study "precise" (concentrated) but not exactly "point" hypothesis, Berger et al. (1994) use frequentist p-values computed from a likelihood conditioned upon a set in which lies the observed dataset, not on the dataset itself, and define a non-decision domain in the BF test procedure. Sellke et al. (2001) advocate calibrating (*rescaling*) the frequentist p-value to relate this new statistic to other test statistics.

As already mentioned in Section 1.1, one can also try to unify the p-value to Bayesian type statistics fully *different* from the BF, to see when frequentist and Bayesian type hypothesis tests can be made equivalent. In particular, when Dempster (1973) proposed to use the PLR, he also mentioned that when testing the mean of a normal distribution, the PLR is equal to the classical frequentist p-value when computed for a uniform prior and with inner parameter $\zeta = 1$. This fundamental result was again emphasized by Aitkin (1997) and Dempster (1997).

Aitkin (1997) asymptotically extended this result to any regular distribution, making use of the asymptotic convergence of a regular distribution towards a normal distribution. For any regular continuous distribution and a smooth prior, the PLR, with $\zeta = 1$, tends asymptotically to the classical p-value. Also, with a nuisance parameter η and still calling θ the tested parameter, he defines LR by $\mathrm{LR}(x, \theta, \eta) = p(x|\theta_0, \eta)/p(x|\theta, \eta)$, in which case under the same conditions as in the previous case the PLR is equal to a p-value. For a normal distribution, when testing the mean and considering the variance as a nuisance parameter, the result is also true for a finite sample.

2.2 New reconciliation result

The sets of conditions found by Dempster (1973) and Aitkin (1997) under which the PLR (with $\zeta = 1$) is equal to a p-value are directly related to the test of the mean of a normal distribution under a uniform prior. The next subsection generalizes this exact finite-sample result under the frame of statistical invariance. As will be discussed at the end of the section, although the technical conditions derived here may be relaxed, it may be difficult to find, at least within the current statistical frame, a fundamentally more general frame of conditions for an equality between the PLR and a p-value to hold.

As presented in current classical textbooks in Bayesian statistics (Berger (1985), Robert (2007)), invariance in statistics arises from the invariant Haar measure defined on some topological group. Throughout this subsection and the related appendices, we will use the notions and results synthesized by Nachbin (1965) and Eaton (1989). The tools necessary to understand the result are introduced in Appendix 1.

In this frame, the PLR (given by an integral over the parameter space Θ) can be reexpressed as an integral over the sample space $\mathcal{X}$, equal to a p-value for $\zeta = 1$. In this subsection x and θ denote random variables or variables of integration according to the context.

First, for clarity, we give the equivalence between the PLR and a frequentist integral under the assumption that the sample space $\mathcal{X}$, the parameter space Θ and the transformations group $\mathcal{G}$ are isomorphic.

Theorem 1. *Call $\mathcal{P}_\Theta = \{p(.|\theta), \theta \in \Theta\}$ a family of probability densities with respect to the Lebesgue measure on $\mathcal{X}$, and call $\mathcal{G}$ a group acting on $\mathcal{X}$. Assume that $\mathcal{P}_\Theta$ is invariant under the action of the group $\mathcal{G}$ on $\mathcal{X}$ and denote by $\bar{g}\theta$ the induced action of the element $g \in \mathcal{G}$ on the element $\theta \in \Theta$. Call H^r and H^l respectively a right and left Haar measure of $\mathcal{G}$ and assume that*

1. *$\mathcal{G}$, $\mathcal{X}$ and Θ are isomorphic.*
2. *The prior measure Π^r is the measure induced by H^r on Θ.*
3. *The measure induced by H^l on $\mathcal{X}$ is absolutely continuous with respect to the Lebesgue measure. Call π^l the corresponding density.*
4. *The marginal density of x is finite, so that the posterior measure Π^r_x on Θ, classically defined by the equation (15), defines the posterior probability $Pr(.|x_0)$.*

Then, the PLR defined by the equation (4) can be reexpressed for any $\zeta > 0$ as the frequentist integral:

$$\mathrm{PLR}(x_0, \zeta) = Pr\left(\frac{p(x_0|\theta_0)}{\pi^l(x_0)} \le \zeta \frac{p(x|\theta_0)}{\pi^l(x)} \,|\, \theta_0 \right) \tag{5}$$

where $x_0 \in \mathcal{X}$ is the observed data and $\theta_0 \in \Theta$ the parameter value under the null hypothesis.

A more general theorem (Theorem 2) derived in a frame which avoids the Lebesgue assumption and may involve more technical conditions is proved in Appendix 2. Theorem 1 is a consequence of Theorem 2 and its proof is given in Appendix 3.

The assumption that $\mathcal{G}$ and $\mathcal{X}$ are isomorphic is easily relaxed by replacing the sample space by the space of a sufficient statistic. Recall that if X is a random variable whose probability distribution is parametrized by θ, $S(X)$ is called a sufficient statistic of θ if the probability distribution of X conditioned upon the random variable $S(X)$ does not depend on θ. Note that according to the Darmois (1935) theorem, among families of probability distributions whose domains do not vary with the parameter being estimated, only in exponential families is there a sufficient statistic whose dimension remains bounded as the sample size increases.

The expression of Theorem 2 is simply extended by replacing X by a sufficient statistic $S(X)$ in the assumptions and by replacing in the frequentist integral the probability density of X by the one of $S(X)$:

Corollary 1. *Call $\mathcal{P}_\Theta = \{p(.|\theta), \theta \in \Theta\}$ a family of probability densities with respect to any measure on $\mathcal{X}$. Call $S(X)$, for $X \in \mathcal{X}$, a sufficient statistic of θ and $\mathcal{P}_{S,\Theta} =$*

$\{p_S(.|\theta), \theta \in \Theta\}$ the family of probability densities of $S(X)$ with respect to the Lebesgue measure on $S(\mathcal{X})$. Call $\mathcal{G}$ a group acting on $S(\mathcal{X})$. Assume that $\mathcal{P}_{S,\Theta}$ is invariant under the action of the group $\mathcal{G}$ on $S(\mathcal{X})$ and denote by $\bar{g}\theta$ the induced action of the element $g \in \mathcal{G}$ on the element $\theta \in \Theta$. Call H^r and H^l respectively any right and left Haar measures of $\mathcal{G}$. Assume that

1. *$\mathcal{G}$, $S(\mathcal{X})$ and Θ are isomorphic.*
2. *The prior measure Π^r is the measure induced by H^r on Θ.*
3. *The measure induced by H^l on $S(\mathcal{X})$ is absolutely continuous with respect to the Lebesgue measure. Call π^l the corresponding density.*
4. *The marginal density of x is finite, so that the posterior measure Π_x^r on Θ, classically defined by the equation (15), defines the posterior probability $Pr(.|x_0)$.*

Then, the PLR defined by the equation (4) can be reexpressed, with $x_0 \in \mathcal{X}$, $\theta_0 \in \Theta$ and $\zeta > 0$, as the frequentist integral:

$$\mathrm{PLR}(x_0, \zeta) = Pr\left(\frac{p_S(S(x_0)|\theta_0)}{\pi^l(S(x_0))} \ \leq \ \zeta \ \frac{p_S(S(x)|\theta_0)}{\pi^l(S(x))} \,|\, \theta_0 \right) \tag{6}$$

where $x_0 \in \mathcal{X}$ is the observed data and $\theta_0 \in \Theta$ the parameter value under the null hypothesis.

The proof follows the proof of Theorem 1 in Appendix 3.

By evaluating $\zeta = 1$ in the result, the PLR with $\zeta = 1$ is easily and finally shown to be equal to a frequentist p-value, where the test statistic is a weighted marginal likelihood of the sufficient statistic $S(x)$.

Corollary 2. *Under the assumptions of Corollary 1, the PLR with inner threshold $\zeta = 1$ is equal to a p-value:*

$$\mathrm{PLR}(x_0, 1) = p_{val}\big(T(x_0)\big) \tag{7}$$

with the test statistic

$$T(x) = \frac{p_S(S(x)|\theta_0)}{\pi^l(S(x))} \,. \tag{8}$$

Corollary 2 can be reexpressed as the fact that under the invariance assumptions, rejecting H_0 when $\mathrm{PLR}(x_0, 1) > p$ is equivalent to rejecting H_0 when $p_{\mathrm{val}}\big(T(x_0)\big) > p$ where the p-value is based on the idea of rejecting H_0 when $T(x_0)$ defined in equation (8) (observed weigthed likelihood under H_0) is not large enough.

2.3 Examples and perspective

Dempster (1973) has shown that the PLR is equal to the classical p-value associated to the test statistic $T(x) = |\bar{x} - \theta_0|$ when testing the mean of a normal family for X with a uniform prior on Θ. Corollary 2 extends this result since the normal family is one of the distributions invariant under translation when testing the location parameter, the uniform prior (i.e. Lebesgue measure) is the measure induced from the right Haar measure associated to translation, and the test statistic $T(.)$ is a monotone function of $p_S(S(.)|\theta_0)\pi^l(S(.))^{-1}$ since the translation (sum) is commutative, so that $\Delta(g) = 1$ for all $g \in \mathcal{G}$ and so π^l is constant.

The result proved here concerns all distributions invariant under some group transformation, under the assumptions that there exists a sufficient statistic and that the sets $\mathcal{G}$, $S(\mathcal{X})$ and Θ are isomorphic. Assume for example that the likelihood p_S has the typical form $p_S(S(x)|\theta) = \theta^{-1} f\big(S(x)\theta^{-1}\big)$. The likelihood is invariant under the scale transformation $g(S(x)) = \alpha \times S(x)$ and the actions on $S(\mathcal{X})$ and Θ are identical. Note that $Uf(U)$ with $U = S(X)\theta^{-1}$ is a pivotal quantity, meaning that its distribution does not depend on θ. The induced prior measure is classically given by $\Pi^r(d\theta) \propto \theta^{-1} d\theta$. Since the multiplication transformation is commutative, the modulus Δ is uniformly equal to 1, so that the test statistic that appears in the p-value (Corollary 2) is simply $T(x) = S(x)\theta_0^{-1} f\big(S(x)\theta_0^{-1}\big)$ where θ_0 is the value of the parameter under H_0. For a more general insight into the relationship between Haar invariance and the Fisher pivotal theory, see Eaton and Sudderth (1999).

Theorem 2 assumes that $\mathcal{G}$, $\mathcal{X}$ and Θ are isomorphic. This assumption is relaxed in Corollaries 1 and 2 where the sample X is replaced by a sufficient statistic $S(X)$: $\mathcal{G}$, $S(\mathcal{X})$ and Θ are assumed to be isomorphic. This trick is one of the two classical dimensionality reduction techniques concerning Haar measures applied to statistical problems and somehow restricts the likelihood to belong to the exponential family from the Darmois theorem. The second trick consists schematically in replacing $S(\mathcal{X})$ by the orbit of $\mathcal{G}$ associated to the observed dataset $O_{x_0} = \{gx_0 \,|\, g \in \mathcal{G}\} \subset \mathcal{X}$. However, the whole set of assumptions that would be involved is more technical, see for example the general assumptions made by Zidek (1969) or Eaton and Sudderth (2002), and not investigated here.

2.4 Connection to other Bayesian and frequentist reconcilations

The result, which concerns hypothesis testing, may be related to the different approaches used to reconcile frequentist and Bayesian point estimation somehow and confidence intervals especially.

Group invariance applied to invariant inference is the classical frame of such unifications. The Fisherian pivotal theory (Fisher (1973, 1st ed.: 1956)) is an important contribution mainly to the "frequentist" side and the right Haar measure to the "Bayesian" side. The reconciliation of the two approaches started with Fraser (1961) and has been deeply studied since then, by Zidek (1969) for example. The most general stage of unification is reached by Eaton and Sudderth (1999). They present the central hypothesis of the Fisherian pivotal theory and show under quite standard assumptions in invariance that this hypothesis leads to a procedure which is identical to the Bayesian invariant procedure when using the prior induced by the right Haar measure. Note that they also show (and in a more general manner by Eaton and Sudderth (2002)) that for a Bayesian invariant inference to be admissible (in the sense that there exists no invariant inference whose mean quadratic error is lower for all θ) it has to be obtained from the right Haar prior.

More concretely, the question related to reconciled probability domains is: "Under what assumptions does the following equality hold?"

$$\begin{aligned} \Pr\big(\theta \in \mathcal{R}(x)\big|x\big) &= \Pr\big(\theta \in \mathcal{R}(x)\big|\theta\big) \\ \text{i.e.} \quad \int_{\{\theta \in \mathcal{R}(x)\}} d\theta\ \pi(\theta|x) &= \int_{\{x\,|\,\theta \in \mathcal{R}(x)\}} dx\ p(x|\theta) \end{aligned} \tag{9}$$

For the equality to hold, each probability needs to be a constant. After the initial work of Fraser (1961), Stein (1965) sketched the first conditions of what would be called later Stein's theorem for invariant domains. The part which is common to the different "Stein's theorems" is the following:

> *If a domain $\mathcal{R}(x) \subset \Theta$ satisfies $\bar{g}\mathcal{R}(x) = \mathcal{R}\big(g(x)\big)$ with $\bar{g}\mathcal{R}(x) = \{\bar{g}\theta\,|\,\theta \in \mathcal{R}(x)\}$, then under [some invariance assumptions],*
>
> $$\begin{aligned} &Pr\big(\theta \in \mathcal{R}(x)\big|x\big) = c \quad \forall x \in \mathcal{X} \quad \textit{(Bayesian probability)} \\ \textit{and } &Pr\big(\theta \in \mathcal{R}(x)\big|\theta\big) = c \quad \forall \theta \in \Theta \quad \textit{(frequentist probability).} \end{aligned}$$

One of the simplest sets of assumptions found since Stein (1965) is the one of Chang and Villegas (1986). It is relatively close to the one used for our results, presented in Section 2.2.

Our result, mainly holding in Theorem 1, is not a consequence of Stein's theorem because the domain $\mathcal{R}(x) \subset \Theta$ is not invariant in our case. $\mathcal{R}(x)$ would be invariant only if θ_0 was invariant under the transformations group $\mathcal{G}$, i.e. if $\bar{g}\theta_0 = \theta_0$ for all $\bar{g}$ (this is equivalent to assuming that H_0 is invariant under $\mathcal{G}$). But in Theorem 2, expressed and proved in Appendix 2 and used in Appendix 3 to prove Theorem 1, ϕ_θ is assumed to be one-to-one for all $\theta \in \Theta$, which implies that $\bar{g}\theta_0 = \theta_0$ is equivalent to $\bar{g} = e$ (identity function). So the domain $\mathcal{R}(x) \subset \Theta$ is not invariant in our case and Stein's theorem does not imply the reconciliation result presented in Section 2.2.

Theorem 1 does not answer the previous question, but rather relaxes the form of the domain and accepts a procedure that varies according to the observed dataset x_0 and the value of the parameter θ_0 under H_0. It answers the question: "Under what assumptions and for what domains $\mathcal{R}$ and $\mathcal{C}$ does the following equality hold?"

$$\int_{\mathcal{R}(x_0,\theta_0) \subset \Theta} d\theta\ \pi(\theta|x_0) = \int_{\mathcal{C}(x_0,\theta_0) \subset \mathcal{X}} dx\ p(x|\theta_0) \tag{10}$$

The domains found take the form

$$\mathcal{R}(x_0,\theta_0) = \{\theta \,|\, p(x_0|\theta_0) \leq p(x_0|\theta)\}$$
$$\mathcal{C}(x_0,\theta_0) = \{x \,|\, p(x_0|\theta_0)f(x_0) \leq p(x|\theta_0)f(x)\}$$

where $f(x)$ is some weighting function, actually given by the inverse of the left prior induced by the underlying group.

3 Concluding general discussion about the PLR

The PLR introduced by Dempster (1973) in the simple vs. composite hypothesis test deserves much attention. It compares the original likelihoods $p(x|\theta_0)$ and $p(x|\theta_1)$ by computing the posterior probability that this usual LR test chooses H_0 or H_1. The PLR is simple, nicely interpretable and coupled with some deep properties. Compared to the classical Bayesian hypothesis tests, first note that unlike the BF, the PLR can be defined even for improper priors, and unlike $\Pr(\mathrm{H}_0|x)$ it does not require the delicate choice of some $\Pr(\mathrm{H}_0)$. This is crucial in practice as well as in fundamental issues like Lindley's paradox.

The PLR also turns out to be a very natural alternative to the BF in many aspects. The PLR first compares (the original likelihoods) and then integrates, whereas the BF first integrates and then compares (the marginal likelihoods). In the simple vs. composite hypothesis test, considering $\mathrm{LR}(x,\theta)$ as a random variable for a fixed x, the PLR is its posterior cumulative distribution (i.e. the probability of a one sided *credible interval*) whereas the BF is its posterior mean *point estimate*. This credible interval vs. point estimate duality between the PLR and the BF also translates in decision theory: Hwang et al. (1992) stressed that $\Pr(\mathrm{H}_0|x)$ does not measure evidence, since this is done only through the likelihood, but measures the accuracy of a test by *estimating* the indicator function $I_{\Theta_0}(\theta)$. Also note that being the measure of a credible interval, the PLR is also a natural hypothesis test tool which connects postdata (i.e. conditioned upon x) hypothesis testing and credible interval inference. This formal equivalence was known to hold for predata inference (a rejection set is equivalent to a confidence interval) and "known" not to hold for postdata inference for usual Bayesian tools (see Lehmann and Romano (2005) and Goutis and Casella (1997)). Tools like the PLR set up this connection.

The connection between the PLR (related to credible interval) and the BF (related to point estimate) has been underlined. Another important connection lies between frequentist and Bayesian type hypothesis tests, namely frequentist p-values and $\Pr(\mathrm{H}_0|x)$ or PLR. This reconciliation quest has been the subject of many debates, including Lindley's paradox in its most simple form (test of the mean of a Gaussian with a uniform prior), which has only been simply reached by the PLR by Dempster (1973). In Section 2.2 we have generalized this reconciliation result to a quite general invariant frame, close to the one used in Stein's theorem, i.e. in a frame under which confidence and credible intervals are equivalent. Note that invariance is also a perspective adopted to develop and evaluate inferences, and in particular to develop new p-values as done recently by

Evans and Jang (2010) for example. For the PLR, standard simple invariance properties directly follow from the simple use of the likelihoods.

To conclude on the contribution of this paper, the equivalence between the PLR and a p-value has been proved in a general invariant frame, which nicely connects to the equivalence between confidence and credible domains. This result may contribute to a better understanding of deep and fundamental issues related to both hypothesis testing and parameter estimation, in both frequentist and Bayesian paradigms.

References

Aitkin, M. (1997). "The calibration of p-values, posterior Bayes factors and the AIC from the posterior distribution of the likelihood." *Statistics and Computing*, 7: 253–261. 942, 944, 945

— (2010). *Statistical inference: an integrated Bayesian / likelihood approach*. Chapman and Hall. 942

Aitkin, M., Boys, R. J., and Chadwick, T. (2005). "Bayesian point null hypothesis testing via the posterior likelihood ratio." *Statistics and Computing*, 25(3): 217–230. 942

Aitkin, M., Liu, C. C., and Chadwick, T. (2009). "Bayesian model comparison and model averaging for small-area estimation." *Annals of Applied Statistics*, 3(1): 199–221. 942

Baskurt, Z. and Evans, M. (2013). "Hypothesis assessment and inequalities for Bayes factors and relative belief ratios." *Bayesian Analysis*, 8,3: 569–590. 941

Berger, J. and Sellke, T. (1987). "Testing a point null hypothesis: the irreconcilability of P values and evidence (with discussion)." *Journal of the American Statistical Association*, 82: 112–139. 944

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer-Verlag, 2nd edition. 945, 955

Berger, J. O., Brown, L., and Wolpert, R. (1994). "A unified conditional frequentist and Bayesian test for fixed and sequential simple hypothesis testing." *Annals of Statistics*, 22(4): 1787–1807. 945

Berger, J. O. and Delampady, M. (1987). "Testing precise hypotheses (with discussion)." *Statistical Science*, 2(3): 317–335. 940, 945

Bernardo, J. (2011). *Bayesian Statistics 9*, chapter Integrated objective Bayesian estimation and hypothesis testing. Oxford University Press. 940

Birnbaum, A. (1962). "On the foundation of statistical inference (with discussion)." *Journal of the American Statistical Association*, 57(298): 269–326. 942

Borges, W. and Stern, J. (2007). "The rules of logic composition for the Bayesian epistemic e-values." *Logic journal of the IGPL*, 15(5–6): 401–420. 942

Casella, G. and Berger, R. L. (1987). "Reconciling Bayesian and frequentist evidence in the one-sided testing problem." *Journal of the American Statistical Association*, 82(397): 106–111. 944

Chang, T. and Villegas, C. (1986). "On a theorem of Stein relating Bayesian and classical inferences in group models." *The Canadian Journal of Statistics*, 14(4): 289–296. 949

Darmois, G. (1935). "Sur les lois de probabilité à estimation exhaustive." *Compte-Rendu de l'Académie des Sciences de Paris*, 200(1265–1266). 946

Dempster, A. P. (1973). "The direct use of likelihood for significance testing." In *Proceedings of Conference on Foundational Questions in Statistical Inference*, 335–354. Aaarhus, Denmark. 941, 942, 944, 945, 948, 950

— (1997). "Commentary on the paper by Murray Aitkin, and on discussion by Mervyn Stone." *Statistics and Computing*, 7(4): 265–269. 942, 945

Eaton, M. (1989). *Group invariance applications in statistics*. Regional Conference Series in Probability and Statistics. 945

— (2007). *Multivariate statistics*. Institute of Mathematical Statistics. 955

Eaton, M. and Sudderth, W. (1999). "Consistency and strong inconsistency of group-invariant predictive inferences." *Bernoulli*, 5(5): 833–854. 948, 949

— (2002). "Group invariant inference and right Haar measure." *Journal of Statistical Planning and Inference*, 103(1–2): 87–99. 948, 949

Evans, M. (1997). "Bayesian inference procedures derived via the concept of relative surprise." *Communications in Statistics*, 26: 1125–1143. 941, 942

Evans, M. and Jang, G. (2010). "Invariant p-values for model checking." *Annals of Statistics*, 38(1): 512–525. 951

Fisher, R. A. (1973, 1st ed.: 1956). *Statistical methods and scientific inference*. Oliver and Boyd, 3rd edition. 948

Fraser, D. A. S. (1961). "The fiducial method and invariance." *Biometrika*, 48(3–4): 261–280. 948, 949

Goutis, C. and Casella, G. (1997). "Relationships between post-data accuracy measures." *Annals of the Institute of Statistical Mathematics*, 49(4): 711–726. 950

Hwang, J., Casella, G., Robert, C., Wells, M., and Farrell, R. (1992). "Estimation of accuracy in testing." *Annals of Statistics*, 20(1): 490–509. 950

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press, 3rd edition. 940, 944

Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses*. Springer, 3rd edition. 940, 950

Lindley, D. (1957). "A statistical paradox." *Biometrika*, 44(1–2): 187–192. 940

Meng, X.-L. (1994). "Posterior predictive p-values." *Annals of Statistics*, 22(3): 1142–1160. 941

Nachbin, L. (1965). *The Haar integral*. Van Nostrand. 945

Newton, M. and Raftery, A. (1994). "Approximate Bayesian inference with the weighted likelihood bootstrap." *Journal of the Royal Statistical Society Series B*, 56(1): 3–48. 942

Neyman, J. (1977). "Frequentist probability and frequentist statistics." *Synthese*, 36: 97–131. 940, 944

Neyman, J. and Pearson, E. (1933). "On the problem of the most efficient tests of statistical hypotheses." *Philosophical Transactions of the Royal Society of London, Series A*, 231: 289–337. 940

Oh, H. and DasGupta, A. (1999). "Comparison of the p-value and posterior probability." *Journal of Statistical Planning and Inference*, 76(1–2): 93–107. 944

O'Hagan, A. (1995). "Fractional Bayes factors for model comparison." *Journal of the Royal Statistical Society*, 57(1): 99–138. 940

Pereira, C. and Stern, J. (1999). "Evidence and credibility: full Bayesian significance test for precise hypotheses." *Entropy*, 1: 104–115. 942

Robert, C. P. (2007). *The Bayesian choice*. Springer, 2nd edition. 940, 945

Robins, J., van der Vaart, A., and Ventura, V. (2000). "Asymptotic distribution of p-values in composite null models." *Journal of the American Statistical Association*, 95(452): 1143–1156. 941

Royall, R. (1997). *Statistical evidence: a likelihood paradigm*. Chapman and Hall / CRC Press. 942

Sellke, T., Bayarri, M. J., and Berger, J. O. (2001). "Calibration of p-values for testing precise null hypotheses." *American Statistician*, 55(1): 62—71. 945

Smith, I. (2010). "Détection d'une source faible : modèles et méthodes statistiques. Application à la détection d'exoplanètes par imagerie directe." Ph.D. thesis, Université de Nice Sophia-Antipolis. 943

Smith, I. and Ferrari, A. (2010). "The posterior distribution of the likelihood ratio as a measure of evidence." In *International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, 391–398. 943

Stein, C. (1965). "Approximation of improper prior measures by prior probability measures." In *Bernoulli, Bayes, Laplace Festschrift*, 217–240. Springer-Verlag. 949

Tsao, C. A. (2006). "A note on Lindley's paradox." *Test*, 15(1): 125–139. 940

Villegas, C. (1981). "Inner statistical inference II." *Annals of Statistics*, 9(4): 768–776. 955

Zidek, J. (1969). "A representation of Bayesian invariant procedures in terms of Haar measure." *Annals of the Institute of Statistical Mathematics*, 21(1): 291–308. 948

Appendix 1: Introduction to invariance in statistics

For a locally compact Hausdorff group $\mathcal{G}$, $K(\mathcal{G})$ denotes the class of all continuous real-valued functions on $\mathcal{G}$ that have compact support. The left invariant Haar measure on $\mathcal{G}$ is defined as a Radon measure H^l such that for all $f \in K(\mathcal{G})$ and all $g_0 \in \mathcal{G}$,

$$\int_{\mathcal{G}} f(g)H^l(dg) = \int_{\mathcal{G}} f(g_0 g)H^l(dg) .$$

The right invariant Haar measure H^r on $\mathcal{G}$ is defined as H^l but replacing $g_0 g$ by $g g_0$. For a given group, both Haar measures exist and are unique up to multiplicative constants.

The (right) modulus Δ of $\mathcal{G}$ is the real positive valued function such that if H^l is a left invariant Haar measure, then for all $f \in K(\mathcal{G})$ and all $g_0 \in \mathcal{G}$,

$$\int f(g g_0^{-1})H^l(dg) = \Delta(g_0)\int f(g)H^l(dg) . \tag{11}$$

From the uniqueness of the Haar measure, Δ does not depend on the choice of H^l and is a continuous function such that for all $g_1, g_2 \in \mathcal{G}$, $\Delta(g_1 g_2) = \Delta(g_1)\Delta(g_2)$, which implies that $\Delta(g^{-1}) = \Delta(g)^{-1}$. Note that for a group $\mathcal{G}$ the set of all right Haar measures is equal to the set of the left Haar measures if and only if Δ is identically equal to 1. This occurs for example when $\mathcal{G}$ is compact or commutative.

Concerning the Haar measures on the group $\mathcal{G}$, the initial definitions and properties imply that if H^l is a left invariant Haar measure on $\mathcal{G}$ and Δ the modulus of $\mathcal{G}$ then for all $f \in K(\mathcal{G})$

$$\int f(g^{-1})H^l(dg) = \int f(g)\Delta(g)^{-1}H^l(dg) . \tag{12}$$

The modulus also enables one to relate right and left invariant Haar measures. From the last property, the measure defined by

$$H^r(dg) = \Delta(g)^{-1}H^l(dg) \tag{13}$$

is a right invariant Haar measure on $\mathcal{G}$. The same way, if H^r is a right invariant Haar measure on $\mathcal{G}$, then the measure defined by $H^l(dg) = \Delta(g)H^r(dg)$ is a left invariant Haar measure.

The Haar measure is applied to statistics through the concept of invariance of a data model under a group of transformations. A parametric family $\mathcal{P}_\Theta = \{p(.|\theta), \theta \in \Theta\}$ of densities with respect to any measure μ on $\mathcal{X}$ is said to be invariant under the transformations group $\mathcal{G}$ if for each $g \in \mathcal{G}$ there exists a unique $\theta^* \in \Theta$ such that if the distribution of X has the density $p(.|\theta) \in \mathcal{P}_\Theta$ then $Y = gX$ has the density $p(.|\theta^*) \in \mathcal{P}_\Theta$. This property defines the action of $\mathcal{G}$ on Θ: θ^* may simply be denoted $\theta^* = \bar{g}\theta$ where $\{\bar{g}, g \in \mathcal{G}\}$ defines a group.

A measure μ on $\mathcal{X}$ is said to be relatively invariant with multiplier χ under the group $\mathcal{G}$ if for all $f \in K(\mathcal{X})$ and $g \in \mathcal{G}$

$$\int f(x)\mu(dx) = \chi(g)\int f(gx)\mu(dx)\,. \tag{14}$$

If we assume that both the family of densities and the measure μ are respectively invariant and relatively invariant, schematically we get $p(x|\theta) = \chi(g)p(gx|g\theta)$ for all $x \in \mathcal{X}, \theta \in \Theta$ and $g \in \mathcal{G}$. For more about the connection between such a multiplier and the Jacobian of the transformation that leads to gx from x, see for example Berger (1985) or Eaton (2007). Note that Theorem 2 could be formulated differently, by defining the invariance of a probability model, but this phrasing is less common than the invariance of a family of probability densities and this would have entailed a longer presentation.

To shorten the preliminaries and without assuming any knowledge about group theory, we will not refer to group properties like transitivity or orbits and will concretely simply assume that Θ and $\mathcal{G}$ are isomorphic. More precisely, we will assume that the transformation $\phi_\theta : \mathcal{G} \mapsto \Theta$ with $\phi_\theta(g) = g\theta$ is one-to-one whatever $\theta \in \Theta$. The right Haar prior on Θ is to be induced from the right Haar measure H^r on $\mathcal{G}$ and the action of $\mathcal{G}$ on Θ. From the frame chosen, the right Haar prior Π_a^r is simply defined by $\Pi_a^r = H^r(\phi_a^{-1})$, with $a \in \Theta$. As shown in Villegas (1981), it turns out that the measure Π_a^r actually does not depend on a. The induced prior is therefore unique for a fixed H^r and noted Π^r. $\Pi^r = H^r(\phi_a^{-1})$ means that for any measurable subset $A \subset \Theta$, $\Pi^r(A) = H^r\big(\phi_a^{-1}A\big)$ with $\phi_a^{-1}A = \big\{\phi_a^{-1}\theta | \theta \in A\big\}$. Note that a subset $A = d\theta$ denotes an infinitesimal subset centered around θ, where θ is implicit. Π^r can be normalized into a probability measure if and only if the group $\mathcal{G}$ is compact, and in this case we can go back to the usual notation $\Pi^r(A) = \Pr(\theta \in A)$ where the measure Π^r is implicit in $\Pr(.)$.

Finally, from the data model density $p(.|\theta)$ and the prior Π^r, the posterior measure Π_x^r on Θ is classically defined by

$$\Pi_x^r(B) = \frac{\int_B p(x|\theta)\Pi^r(d\theta)}{m(x)} \qquad \text{for all} \quad B \subset \Theta \tag{15}$$

$$\text{with} \quad m(x) = \int p(x|\theta)\Pi^r(d\theta)$$

where the marginal $m(x)$ density of x is always assumed to be finite, so that Π_x^r defines a probability measure even if Π^r does not. Then the posterior probability of an event is denoted by $\Pr(.|x)$, meaning $\Pr(\theta \in B|x) = \Pi_x^r(B)$.

Appendix 2: General theorem and its proof

Theorem 2. *Call $\mathcal{P}_\Theta = \{p(.|\theta), \theta \in \Theta\}$ a family of probability densities with respect to a measure μ^r on $\mathcal{X}$, specified later, and call $\mathcal{G}$ a group acting on $\mathcal{X}$. Assume that $\mathcal{P}_\Theta$ is invariant under the action of the group $\mathcal{G}$ on $\mathcal{X}$ and denote by $\bar{g}\theta$ the induced action of the element $g \in \mathcal{G}$ on the element $\theta \in \Theta$. Call H^r any right Haar measure of $\mathcal{G}$ and define the transformations ϕ_θ (for $\theta \in \Theta$) and ϕ_x (for $x \in \mathcal{X}$) by*

$$\begin{aligned} \phi_\theta : \bar{\mathcal{G}} &\mapsto \Theta \\ \bar{g} &\mapsto \bar{g}\theta \end{aligned} \qquad\qquad \begin{aligned} \phi_x : \mathcal{G} &\mapsto \mathcal{X} \\ g &\mapsto gx \end{aligned} \tag{16}$$

Assume that

1. *ϕ_θ is one-to-one for all $\theta \in \Theta$ and ϕ_x is one-to-one for all $x \in \mathcal{X}$.*

2. *The prior measure Π^r on Θ is the measure induced by H^r via ϕ_θ and the measure μ^r on $\mathcal{X}$ is the measure induced by H^r via ϕ_x: $\Pi^r = H^r(\phi_\theta^{-1})$ and $\mu^r = H^r(\phi_x^{-1})$.*

3. *The marginal density of x is finite, so that the posterior measure Π_x^r on Θ, classically defined by the equation (15), defines the posterior probability $Pr(.|x_0)$.*

Then, the PLR defined by the equation (4) can be reexpressed, for any $\zeta > 0$ and any *$c \in \mathcal{X}$, as the frequentist integral:*

$$\mathrm{PLR}(x_0, \zeta) = Pr\Big(\ p(x_0|\theta_0)\Delta\big(\phi_{x_0}^{-1}c\big)\ \le\ \zeta\ p(x|\theta_0)\Delta\big(\phi_x^{-1}c\big)\,|\,\theta_0\Big) \tag{17}$$

where Δ is the modulus of the group $\mathcal{G}$, as defined in equation (11), and in practice $x_0 \in \mathcal{X}$ is the observed data and $\theta_0 \in \Theta$ the parameter value under the null hypothesis.

Proof: Note that as seen in the previous appendix, the measures μ and Π^r defined in Theorem 2 do not depend on the choice of $\theta \in \Theta$ and $x \in \mathcal{X}$ in the functions ϕ_θ and ϕ_x. In order to clarify the proof, we note a instead of x and b instead of θ in the following. We shall make use of the following lemma:

Lemma 1. *The measures μ on $\mathcal{X}$ and Π^r on Θ induced above by the right Haar measure H^r on $\mathcal{G}$ are relatively invariant with modulus Δ^{-1}.*

Proof:

$$\begin{aligned}
\int f(g_0 x)\mu(dx) &= \int f(g_0 x) H^r \phi_a^{-1}(dx) \quad \text{(Def. of } \mu \text{ in the Cond. of Th. 2)} \\
&= \int f\big(g_0 \phi_a g\big) H^r(dg) \quad \text{(transformation } g = \phi_a^{-1} x) \\
&= \int f\big(g_0 g a\big) \Delta(g)^{-1} H^l(dg) \quad \text{(Def. of } \phi_a \text{ and prop. eq. (13))} \\
&= \Delta(g_0) \int f\big(g_0 g a\big) \Delta(g_0 g)^{-1} H^l(dg) \quad \text{(Multiplicity prop. of } \Delta) \\
&= \Delta(g_0) \int f(ga) \Delta(g)^{-1} H^l(dg) \quad (H^l \text{ left invariant}) \\
&= \Delta(g_0) \int f(x)\mu(dx) \quad \text{(previous computation made in reverse order)}
\end{aligned}$$

This also implies that a Haar prior induced as in Theorem 2, i.e. from a right invariant Haar measure on $\mathcal{G}$, is relatively invariant.

$$\begin{aligned}
\mathrm{PLR}(x_0, \zeta) &= \Pr\Big(p(x_0|\theta_0) \le \zeta p(x_0|\theta) \;\Big|\; x_0\Big) \\
&= \frac{1}{m(x_0)} \int_{\left\{\theta \,|\, p(x_0|\theta_0) \le \zeta p(x_0|\theta)\right\}} p(x_0|\theta) \Pi^r(d\theta) \qquad (18) \\
&= \frac{1}{m(x_0)} \int_{\left\{\theta \,|\, p(x_0|\theta_0) \le \zeta p(x_0|\theta)\right\}} p(x_0|\theta) H^r\big(\phi_b^{-1}(d\theta)\big) \quad \text{(Def. } \Pi^r \text{ in Th. 2)} \\
&= \frac{1}{m(x_0)} \int_{\left\{g \,|\, p(x_0|\theta_0) \le \zeta p(x_0|\phi_b g)\right\}} p(x_0|\phi_b g) H^r(dg) \quad (g = \phi_b^{-1}\theta) \\
&= \frac{1}{m(x_0)} \int_{\left\{g \,|\, p(x_0|\theta_0) \le \zeta p(x_0|gb)\right\}} p(x_0|gb) H^r(dg) \quad \text{(Def. } \phi_\theta \text{ eq. (16))} \\
&= \frac{1}{m(x_0)} \int_{\left\{g \,|\, p(x_0|\theta_0) \le \zeta p(x_0|gb)\right\}} p(x_0|gb) \Delta(g)^{-1} H^l(dg) \quad \text{(Prop. eq. (13))} \\
&= \frac{1}{m(x_0)} \int_{\left\{g \,|\, p(x_0|\theta_0) \le \zeta p(x_0|g^{-1}b)\right\}} p\big(x_0|g^{-1}b\big) H^l(dg) \quad \text{(Prop. eq. (12))}
\end{aligned}$$

But according to lemma 1, μ is relatively invariant with modulus Δ^{-1}. Since the density family is invariant,

$$p(x|\theta) = \Delta(g)^{-1} p(gx|g\theta) \quad \text{for all } x \in \mathcal{X}, \theta \in \Theta, g \in \mathcal{G}$$

i.e.

$$p(x|g^{-1}\theta') = \Delta(g)^{-1} p(gx|\theta') \quad \text{for all } x \in \mathcal{X}, \theta' \in \Theta, g \in \mathcal{G}$$

Then,

$$\begin{aligned}
\mathrm{PLR}(x_0,\zeta) &= \frac{1}{m(x_0)} \int_{\left\{g:\,|\,p(x_0|\theta_0)\le \zeta p(gx_0|b)\Delta(g)^{-1}\right\}} \Delta(g)^{-1} p(gx_0|b) H^l(dg) \\
&= \frac{1}{m(x_0)} \int_{\left\{g\,|\,p(x_0|\theta_0)\le \zeta p(gx_0|b)\Delta(g)^{-1}\right\}} p(gx_0|b) H^r(dg) \quad \text{(Prop. eq. (13))} \\
&= \frac{1}{m(x_0)} \int_{\left\{g\,|\,p(x_0|\theta_0)\le \zeta p(gx_0|b)\Delta(g)^{-1}\right\}} p(gx_0|b) \mu\big(\phi_a(dg)\big) \quad \text{(Def. } \mu) \ .
\end{aligned}$$

It can be noticed that the equation (18) depends neither on $a \in \mathcal{X}$ nor on $b \in \Theta$. Choose now for simplicity $a = x_0$. Then, making the transformation $x = \phi_{x_0} g = gx_0$,

$$\mathrm{PLR}(x_0,\zeta) = \frac{1}{m(x_0)} \int_{\left\{x\,|\,p(x_0|\theta_0)\le \zeta p(x|b)\Delta\left(\phi_{x_0}^{-1}x\right)^{-1}\right\}} p(x|b)\mu(dx) \ .$$

By a similar computation we get the expression of the marginal density of X evaluated at x_0:

$$m(x_0) = \int p(x_0|\theta)\Pi^r(d\theta) = \int p(x|b)\mu(dx) = 1 \ .$$

The marginal density of X is constant, the same way the frequentist risk of an invariant estimator does not depend on θ. So

$$\mathrm{PLR}(x_0,\zeta) = \int_{\left\{x\,|\,p(x_0|\theta_0)\le \zeta p(x|b)\Delta\left(\phi_{x_0}^{-1}x\right)^{-1}\right\}} p(x|b)\mu(dx) \ .$$

In order to get a form closer to a p-value, we choose from now $b = \theta_0$ and note that for *any* $c \in \mathcal{X}$,

$$\Delta\big(\phi_{x_0}^{-1}x\big) = \frac{\Delta\big(\phi_c^{-1}x\big)}{\Delta\big(\phi_c^{-1}x_0\big)} \tag{19}$$

because if we note

$$\begin{aligned}
g &= \phi_{x_0}^{-1}x \\
g_1 &= \phi_c^{-1}x \\
g_2 &= \phi_c^{-1}x_0
\end{aligned}$$

then on one side $gx_0 = x$ and on the other $g_1(g_2^{-1}x_0) = g_1 c = x$ so that

$$\begin{aligned}
gx_0 &= (g_1 g_2^{-1})x_0 \\
\text{so} \quad \phi_{x_0} g &= \phi_{x_0}(g_1 g_2^{-1}) \\
\text{so} \quad g &= g_1 g_2^{-1} \quad (\phi_a \text{ is one-to-one}) \\
\text{so} \quad \Delta(g) &= \frac{\Delta(g_1)}{\Delta(g_2)} \quad \text{(Prop. of } \Delta) \ .
\end{aligned}$$

Finally, for any $c \in \mathcal{X}$

$$\mathrm{PLR}(x_0, \zeta) = \int_{\left\{x \mid \frac{p(x_0|\theta_0)}{\Delta(\phi_c^{-1} x_0)} \leq \zeta \frac{p(x|\theta_0)}{\Delta(\phi_c^{-1} x)}\right\}} p(x|\theta_0)\mu(dx)\ . \tag{20}$$

It is also interesting to note that

$$\phi_a^{-1} b = (\phi_b^{-1} a)^{-1} \tag{21}$$
$$\text{since } g = \phi_a^{-1} b \Rightarrow ga = b \Rightarrow a = g^{-1} b \Rightarrow g^{-1} = \phi_b^{-1} a$$

so that the same way we have

$$\begin{aligned}\mathrm{PLR}(x_0, \zeta) &= \int_{\left\{x \mid p(x_0|\theta_0)\Delta\left(\phi_{x_0}^{-1} c\right) \leq \zeta p(x|\theta_0)\Delta\left(\phi_x^{-1} c\right)\right\}} p(x|\theta_0)\mu(dx) \\ &= \Pr\Big(p(x_0|\theta_0)\Delta\big(\phi_{x_0}^{-1} c\big) \leq \zeta p(x|\theta_0)\Delta\big(\phi_x^{-1} c\big)\Big|\theta_0\Big)\ .\end{aligned}$$

This ends the proof. □

Appendix 3: Proof of Theorem 1 and Corollaries 1, 2

Theorem 1 is a corollary of Theorem 2 presented and proved in the previous appendix: Theorem 2 can be reexpressed more simply by assuming that the likelihood family and the induced Haar measures are absolutely continuous with respect to the Lebesgue measure.

Proof of Theorem 1: The proof only consists of reexpressing the domains of integration because the integrands expression are not functions of the use of the decomposition of the measures over some other measures (μ or Lesbesgue). The proof even actually only consists of reexpressing the domain of integration of the p-value because the domain of integration of the PLR does not depend on the density over $\mathcal{X}$ used since the domain of integration is a subset of Θ, not $\mathcal{X}$.

If we denote by $p^\mu(.|\theta)$ the density with respect to the induced Haar measure μ^r and by $p(.|\theta)$ the density with respect to the Lebesgue measure, we have by definition

$$P(dx|\theta) = p^\mu(x|\theta)\mu^r(dx) = p^\mu(x|\theta)\pi^r(x)dx \quad \text{and} \quad P(dx|\theta) = p(x|\theta)dx$$
$$\text{and so} \quad p^\mu(x|\theta) = \frac{p(x|\theta)}{\pi^r(x)}\ .$$

On the other side the modulus Δ can also be reexpressed as a function of the induced prior densities $\pi^l(x)$ and $\pi^r(x)$. From equations (21) and (13),

$$\Delta\left(\phi_x^{-1} c\right) = \Delta\left(\phi_c^{-1} x\right)^{-1} = \frac{H^l\left(d\phi_c^{-1} x\right)}{H^r\left(d\phi_c^{-1} x\right)} = \frac{\mu^r(dx)}{\mu^l(dx)} = \frac{\pi^r(x)}{\pi^l(x)}\ .$$

Combining these two results we get

$$p^{\mu}(x|\theta)\Delta\left(\phi_x^{-1}c\right) = \frac{p(x|\theta)}{\pi^l(x)} \,.$$

□

Proof of Corollary 1:

$$\begin{aligned}
\mathrm{PLR}(x_0,\zeta) &= \Pr\Big((p_{X|\theta_0}(x_0) \;\leq\; \zeta\; p_{X|\theta}(x_0)\;\Big|\; x_0\Big) \\
&= \Pr\Big(p_{X|S(X)}(x_0|S(x_0))\; p_{S(X)|\theta_0}(S(x_0)) \\
&\qquad\qquad \leq\; \zeta\; p_{X|S(X)}(x_0|S(x_0))\; p_{S(X)|\theta}(S(x_0))\;\Big|\; x_0\Big)
\end{aligned}$$

because since $S(x)$ is a function of x, $p_{X|\theta}(x) = p_{X,S(X)|\theta}\big(x, S(x)\big)$ and since in addition $S(X)$ is a sufficient statistic of X,

$$\begin{aligned}
p_{X|\theta}(x) &= p_{X|S(X),\theta}\big(x|S(x)\big)\; p_{S(X)|\theta}\big(S(x)\big) \\
&= p_{X|S(X)}\big(x|S(x)\big)\;\; p_{S(X)|\theta}\big(S(x)\big)\,.
\end{aligned} \tag{22}$$

Simplifying the densities which do not depend on θ,

$$\begin{aligned}
\mathrm{PLR}(x_0,\zeta) &= \Pr\Big(p_{S(X)|\theta_0}\big(S(x_0)\big) \;\leq\; \zeta\; p_{S(X)|\theta_0}\big(S(x_0)\big)\;\Big|\; S(x_0)\Big) \\
&= \Pr\Big(p_{S(X)|\theta_0}\big(S(x_0)\big)(\pi^l(S(x_0)))^{-1} \\
&\qquad\qquad \leq\; \zeta\; p_{S(X)|\theta_0}\big(S(x)\big)(\pi^l(S(x)))^{-1}\Big|\theta_0\Big) \qquad (\text{Th. } 1)\;\;.
\end{aligned}$$

□

Proof of Corollary 2: First reexpress the PLR under the conditions of Theorem 1 by using a cumulative distribution. Note $T(x)$ the statistic:

$$T(x) = p_{S(X)|\theta_0}\big(S(x)\big)(\pi^l(S(x)))^{-1}\,.$$

Seen as a random variable, the dataset x induces the random variable $T(X)$ the same way the statistic $S(x)$ induced $S(X)$. Note $F_{T(X)|\theta_0}$ the cumulative distribution of $T(X)$ under the null hypothesis:

$$F_{T(X)|\theta_0}(\zeta) = \Pr\big(T(x) \leq \zeta\big|\theta_0\big)\,.$$

Starting from the theorem 1, the PLR can be reexpressed as

$$\mathrm{PLR}(x_0,\zeta) = 1 - F_{T(X)|\theta_0}\big(\zeta^{-1}T(x_0)\big)\,.$$

In particular, for a threshold $\zeta = 1$, one can directly notice that the PLR is equal to the *p-value* defined for the GLR by equation (2), but now instead associated to the test statistic $T(x)$.

Also note that the frequentist test corresponding to the PLR is then given, for any threshold $\lambda > 0$, by

$$\text{Reject } H_0 \text{ if} \qquad p_S\big(S(x)|\theta_0\big)\ (\pi^l(S(x)))^{-1}\ \leq\ \lambda\,.$$

□