A new criterion for prior probabilities

Cristofaro Rodolfo De, Bruno Lecoutre

A New Criterion for Prior Probabilities

Rodolfo de Cristofaro [a,*] Bruno Lecoutre [b]

[a] *Università degli Studi di Firenze (Italy).*

[b] *ERIS, Laboratoire de Mathématiques Raphaël Salem, CNRS-Université de Rouen Avenue de l'Université, BP 12, 76801 Saint-Etienne-du-Rouvray (France) bruno.lecoutre@univ-rouen.fr.*

Abstract

Howson and Urbach (1996) wrote a carefully structured book supporting the Bayesian view of scientific reasoning, which includes an unfavorable judgment about the so-called objective Bayesian inference. In this paper, the theses of the book are investigated from Carnap's analytical viewpoint in the light of a new formulation of the Principle of Indifference. In particular, the paper contests the thesis according to which no theory can adequately represent 'ignorance' between alternatives. Beginning from the new formulation of the principle, a criterion for the choice of an objective prior is suggested in the paper together with an illustration for the case of Binomial sampling. In particular, it will be shown that the new prior provides better frequentist properties than the Jeffreys interval.

Key words: Data translated likelihood, Frequentist properties, Inductive logic, Jeffreys-rule, Likelihood principle, Objective Bayesian analysis, Principle of indifference, Stopping rule

1 Introduction

Howson and Urbach's book titled "Scientific Reasoning: The Bayesian Approach" (1996) is a carefully argued exposition and defense of the Bayesian view of scientific reasoning. In particular, according to Howson and Urbach,

* On october 2007, Rodolfo de Cristofaro submitted for publication a preliminary version of this paper entitled "Remarks regarding the choice of an objective prior distribution." He had heart disease that led to his death on July 21, 2008. The current text is a revised, enlarged version, which was written in collaboration with the second author before that date. Only minor corrections were made after that date, and thus are the sole responsibility of the second author.

alternative methods of inference, especially those connected with significance testing and estimation, are really quite unsuccessful and, despite their influence among scientists, their pre-eminences are undeserved. What is more, subjectivity in the Bayesian approach is, first of all, minimal and, secondly, exactly right. The ideal of total objectivity is unattainable and alternative approaches to scientific reasoning, which pose as guardian of that ideal, in fact violate it at every turn; virtually no other method can be applied without a generous helping of personal judgment and arbitrary assumption.

We agree with the arguments advanced by the authors concerning inductive reasoning; arguments, which are in line with Rudolf Carnap's thinking on inductive logic. There is, however, a difference. It involves the analytical character of the solution to be given to the problem of statistical induction.

As Carnap (1962, p. 518) wondered, "why did statisticians spend so much effort in developing methods of inference independent of the probability axioms? It seems clear that the main reason was purely negative; it was the dissatisfaction with the principle of indifference (or insufficient reason). If we should find a degree of confirmation which does not lead to the unacceptable consequences of the principle of indifference, then the main reason for developing independent methods of estimation and testing would vanish. Then it would seem more natural to take the degree of confirmation as the basic concept for all of inductive statistics."

Today, are we able to say whether the unacceptable consequences of the principle of indifference have been avoided? It is hard to answer this question in the affirmative. Should the prior information be held to be irrelevant, a probability distribution should exist to be assigned to given hypotheses on which different individuals agree.

In other words, in order to control induction *analytically*, one would have to know how to assign the probabilities of the various hypotheses in case prior information is held to be irrelevant. Moreover, this should allow evaluating the subjective component due to other information (the information preceding the current experimentation).

Another point, which deserves to be investigated, has to do with the likelihood principle (LP). LP concerns foundations of statistical inference and it is often invoked in arguments about correct statistical reasoning. Let $f(x \mid \theta)$ be a conditional distribution for X given the unknown parameter $\theta \in \Theta$ (the set of possible values of θ). According to the LP, in the inference about θ, after $X = x$ is observed, all relevant experimental information is contained in the likelihood function for the observed x. As an operative implication of this principle, if different designs produce proportional likelihood functions, one should make an identical inference about a parameter θ from the data, x, irrespective of the particular design that yields x.

There are several counter-examples, and/or paradoxical consequences to the LP. What is more, some methods of conventional statistics are not consistent with the LP. For reference, see Severini (2000, pp. 79 ff.). The authors who do not accept these consequences reject the utilization of Bayes formula, and solve statistical inference by simply examining the likelihood function under a weak version of the LP. On this subject, one may see Cox and Hinkley (1974, p. 39).

If we accept the likelihood principle in its weak version, then the approach deriving from it is objective (or would at least allow one to assess the experimental information in an objective manner). Nevertheless, as we were able to show in a previous article (cf. de Cristofaro, 2004), the LP (both in its strong and weak version) is questionable. In fact, contrary to a widely held opinion, it is not a consequence of Bayes theorem.

To be honest, the differences, compared to the book by Howson and Urbach, do not involve the Bayesian approach — with which we basically agree — but the judgment about the choice of an objective prior distribution, and, more in general, the foundations of the so-called 'Objective Bayesian Analysis'.

2 On the irrelevance of stopping rule

Howson and Urbach (1996) do not speak explicitly of the likelihood principle, although they do deal with the issue when referring to the stopping rule (the rule that dictates when the trial should terminate).

The discussion is related to the example of an experiment that is designed to elicit a coin's physical probabilities of landing heads and tails. And the conclusion is as follows (p. 365): it does not matter whether the experimenter intended to stop after n tosses of the coin or after r heads appeared in a sample; the inference about θ [the probability of landing heads] is exactly the same in both cases.

The LP is based on an assumption held to be obvious and therefore not verified, which is the assignment of the prior probabilities about the parameter θ independently from the manner in which the trials were conducted: the so-called sampling rule, experiment, process of data generation, or — as we will call it — design, d.

Quoting Lindley's words (2000), only the realized actual observation x is relevant: "This is the likelihood principle according to which values of x, other than that observed, play no role in inference." Nevertheless, the correct version of Bayes formula shows that, not only the data, but also d is relevant in inference.

Suppose that x is the observed value of a random variable X, whose probability distribution $p(x \mid \theta, e)$ depends on θ and e. Suppose also that θ itself has a probability distribution $p(\theta \mid e)$ conditional on e, called the prior distribution. Then, given x and e, the posterior distribution of θ, $p(\theta \mid x, e)$, is

$$p(\theta \mid x, e) \propto p(\theta \mid e) p(x \mid \theta, e). \tag{1}$$

The statement of (1) is the expression of Bayes formula in accordance with the Carnap's philosophy.

Indeed, e comprises not only the beliefs of the experimenter before the experiment is performed, e^*, but also the piece of information about d. That is, $e = (e^*, d)$.

In particular, x is determined by a particular design d with a given θ. Hence, $p(x \mid \theta, e)$ is not defined without a reference to d. Thus, the probability of x successes on n trials is different according to the supposed process of data generation: direct or inverse sampling, hypergeometric scheme, Markov process, and so on.

The conditional assertion about d is usually omitted when it is clear from the context which design has been chosen. But if two people assume the same prior with reference to different designs, it becomes important to spell out the designs each used. In fact, the correct expression of Bayes formula shows that the prior $p(\theta \mid e^*, d)$ depends on d.

To conclude, the core of the LP is that, given x, d is irrelevant under the same likelihood. On the contrary, whatever the data may be, the evidence about d may affect the prior, and, consequently, the result of inference.

The omission of any reference to evidence has been the cause of unnecessary debates and unpleasant consequences. The very heart of the matter is the ancillary knowledge of how the data were collected. In fact, as we saw, d is a full part of e. It may play a basic role in inference.

In this regards, the often quoted sentence by Edwards, Lindman, and Savage (1963) should be revised: "The likelihood principle emphasized in Bayesian statistics implies, among other things, that the rules governing when data collection stops are irrelevant to data interpretation. It is entirely appropriate to collect data until a point has been proven or disproved, or until the data collector runs out of time, money, or patience."

This recommended practice appears as a pill hard to digest for experimenters willing to adopt Bayesian methods. Designs are open to unscrupulous manipulation if the experimenter is allowed to ignore the rules governing the data sampling and to choose the stopping point irrespectively of the prior.

In reality, Bayes formula (in its correct expression) does not allow us to choose at discretion the data, to ignore the rules governing their collection, or to stop the sampling irrespective of the prior.

We can find in Severini (2000, p. 79) a remark essentially correct: "information beyond that provided by the likelihood function is necessary for proper statistical inference". But the right answer to this remark is not the weak version of LP. It is the evidence about d and its effect on the prior.

Applying this idea, Bunouf and Lecoutre (2006, 2008, 2010) developed Jeffreys-type priors derived from likelihood augmented with the design information in multistage designs. They showed that the use of such priors corrects the posteriors from the stopping rule bias.

For a more detailed analysis regarding the foundations of LP, see de Cristofaro (2004).

3 The new principle of indifference

The principle of indifference is a rule for assigning probabilities under 'ignorance'. It was called "of indifference" by John M. Keynes, who was careful to note that it applies only when there is no knowledge indicating unequal probabilities. Anyway, we are not in a situation of total ignorance, since we know the design d that is going to generate the data. Thus, if we know that d is able to favor some hypothesis, then indifference principle does not apply.

Quoting Howson and Urbach (1996, p. 363), a fact is relevant to a set of hypotheses when knowing it makes a difference to one's appraisal of those hypotheses. In this connection, is it relevant for the assignment of an equal probability to each face of a die whether it is or not biased or whether its casting is or not fair? I think so. In the same way, apart from other information, in order to assign the same probability to every admissible hypothesis, the design d should be 'fair' or 'impartial', in the sense of ensuring the same support to all hypotheses.

This reasoning leads us to the following definition: If, under d, all hypotheses are equally supported whatever the data may be, we shall say that d is *impartial*.

That being stated, the new *Principle of Indifference* is as follows (cf. de Cristofaro, 2008):

> Given the set of all admissible hypotheses H, let h denote any one element of the partition of H and let d denote the projected design, then we are allowed to assign the same probability to every h if (i) prior information is considered to be irrelevant, and (ii) d is impartial.

In statistics, in order to ensure the impartiality of d towards the parameter θ, it is sufficient that the superior extreme ordinate of the posterior $p(\theta \mid x, d)$, for a possible x, is situated on the same level of any other curve of the posterior obtainable from d (*superior profile criterion*).

In plain language, if d is genuinely impartial to θ, then all possible values of θ should be equally supported whatever the data may be. Likewise, the superior extreme ordinate of the posterior curves graphed for all possible data, x, should be constant.

Let $\ell'(\theta \mid t)$ be the standardized likelihood of θ given a possible observation of the sufficient statistic t about θ (whose integral with respect to $\theta \in \Theta$ is 1). Apart from a proportionality constant that does not depend on t or θ, a means whereby profile criterion can be put to the test consists in plotting the function

$$h(\theta) = \sup_{t} \ell'(\theta \mid t), \tag{2}$$

in order to see whether it is or not constant. Of course, the assumption of a uniform prior for θ is justified in the affirmative.

For instance, the Binomial mean θ is far from being impartial. In case of $n = 24$ trials and x successes, we have:

$$h(\theta) = \sup_{x=0,\ldots,24} \frac{25!}{x!(24-x)!} \theta^x (1-\theta)^{24-x}. \tag{3}$$

This function is U-shaped, and, therefore, the assumption of a uniform density prior for θ is not justified. (cf. the standardized likelihood curves in Figure 1).

A similar criterion about the choice of an objective prior was introduced by Fisher in the year 1922 and it was worked out by Box and Tiao (1973; for reference to Fisher, see p. 35). According to these authors, the prior distribution for a parameter, let us say θ, is assumed to be locally uniform if different sets of data translate the likelihood curve on the θ-axis, leaving it unchanged in shape and spread (that is, the data only serve to change the location of the likelihood). On the other hand, if θ is not *data translated* in this sense, then Box and Tiao suggest expressing the parameter in terms of a new metric $\phi = \phi(\theta)$, so that the corresponding likelihood is 'data translated'.

As we can see from the examples and figures shown by Box and Tiao (1973, pp. 27-39) if the parameter θ is data translated (and the design d is impartial to θ), then the superior extreme ordinate of every possible curve of the standardized likelihood for θ is situated on the same level of any other curve obtainable from d.

For illustration, suppose $\mathbf{x}' = (x_1, ..., x_n)$ is a random sample from a Normal distribution $N(\mu, \sigma^2)$, where σ is supposed known. The likelihood function of μ is

$$\ell(\mu \mid \mathbf{x}) \propto \exp\Big[-\frac{n}{2\sigma^2}(m-\mu)^2\Big], \tag{4}$$

where m is a possible determination of the average of observations. This function is represented by a Normal curve with its maximum value that remains constant for all possible determinations of m. In particular, when $m = \mu$, (4) is proportional to a constant. Thus, the design is impartial to μ, and the density prior for μ can be assumed locally uniform.

Now, it could happen that the quantity of interest was not μ but its reciprocal $\gamma = \mu^{-1}$. In this case, the posterior or standardized likelihood for γ is

$$\ell'(\gamma \mid \sigma, m) \propto \gamma^{-2} \exp\Big[-\frac{n}{2\sigma^2}(m-\gamma^{-1})^2\Big], \tag{5}$$

which is a curve with its maximum ordinate proportional to γ^{-2}. In particular, when $m = \gamma^{-1}$, (5) is proportional to γ^{-2}.

On the other hand,

$$p(\gamma \mid \sigma) = p(\mu \mid \sigma)\Big|\frac{d\mu}{d\gamma}\Big| = p(\mu \mid \sigma)\mu^2 \propto \gamma^{-2}. \tag{6}$$

As we saw with reference to (5), the density prior for γ (besides being proportional to the Jacobian of the transformation from μ to γ) is proportional to the superior profile of the posterior (or standardized likelihood) curves for γ, with reference to all possible determinations of m from the intended design. Namely,

$$p(\gamma \mid \sigma) \propto \sup_m \ell'(\gamma \mid \sigma, m) = \gamma^{-2}. \tag{7}$$

Notice that the reference to 'standardized likelihood' is due to Box and Tiao

(1973), where we can find some unremarked illustrations about the property of the prior we have just mentioned (cf. pp. 27, 30, 35, 39).

More in general, if we do a one-to-one transformation of the parameter θ concerning an impartial design in terms of a new metric $\phi = \phi(\theta)$, then the prior

$$p(\phi) \propto \sup_{\hat{\phi}} p(\phi \mid \hat{\phi}), \tag{8}$$

where $p(\phi \mid \hat{\phi})$ is the posterior for ϕ, given a possible observation of the sufficient statistic $\hat{\phi}$ about ϕ.

We can see the argument from another viewpoint: if (8) holds, then exists a transformation $\theta = \theta(\phi)$ that makes the design impartial with respect to θ. The profile criterion, we suggested with reference to the uniform distribution, is a particular case of a more general rule, given by (8).

According to this rule, the prior probability for the parameter ϕ conditional to d, $p(\phi \mid d)$, is proportional to the superior profile of corresponding posterior curves, $p(\phi \mid x, d)$, considered for all possible x obtainable from d. A rule we can apply to any prior.

As an example, the density prior for Binomial mean θ suggested by Jeffreys is

$$p(\theta) \propto [\theta(1-\theta)]^{-1/2}. \tag{9}$$

This prior is approximately proportional to the superior profile of standardized likelihood curves for given n

$$\ell'(\theta \,|\, x) = \frac{1}{B(x+1, n-x+1)}\, \theta^x (1-\theta)^{n-x}, \tag{10}$$

where $B(\alpha, \beta)$ is the complete Beta function. Note that $\ell'(\theta \,|\, x)$ is the density of a Beta distribution of parameters $x+1$ and $n-x+1$, which is the posterior distribution for θ associated with an uniform prior.

In the case of $n = 24$ trials, Figure 1 shows the curves of standardized likelihood for $x = 1$, 3, 7, 12, 17, 21, 23 successes, together with the prior of (9). As we can see, (9) is approximately (with a very good approximation) proportional to the corresponding maximum of the curves shown in Figure 1.

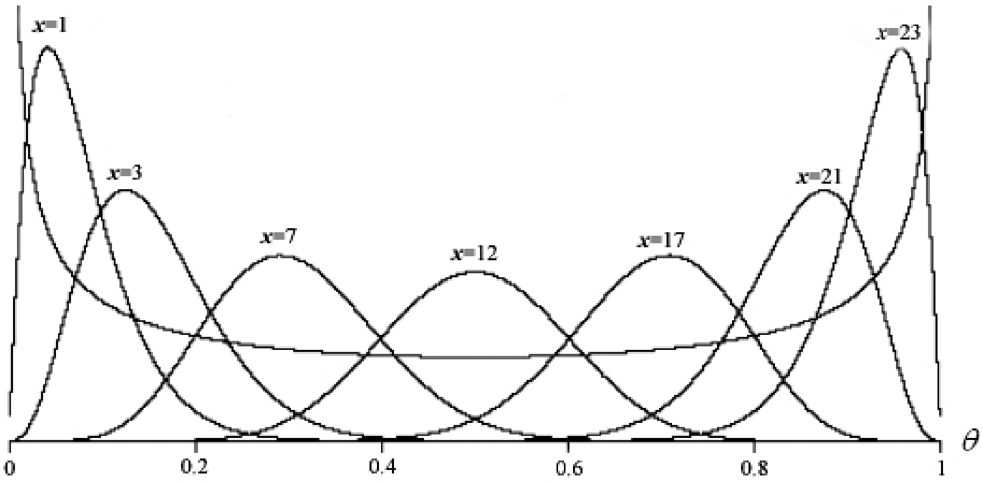

Figure 1. Standardized likelihood curves for Binomial mean θ and Jeffreys prior distribution.

A similar approximation holds between (9) and

$$\sup_x \frac{1}{B(x+\frac{1}{2}, n-x+\frac{1}{2})}\, \theta^{x-\frac{1}{2}}(1-\theta)^{n-x-\frac{1}{2}}. \tag{11}$$

in which we can recognize the density of a Beta distribution of parameters $x+\frac{1}{2}$ and $n-x+\frac{1}{2}$, which is the posterior distribution for θ associated with the Jeffreys prior $\beta(\frac{1}{2},\frac{1}{2})$.

Because of this approximation, the transformation $\phi = sin^{-1}\sqrt{\theta}$ does not make the Binomial experiment exactly impartial. As we can see in Figure 2, the superior profile of the standardized likelihood curves for ϕ is not constant, although it is nearly so. If necessary, a prior may be assumed which is approximately uniform. But, strictly speaking, the design is not impartial.

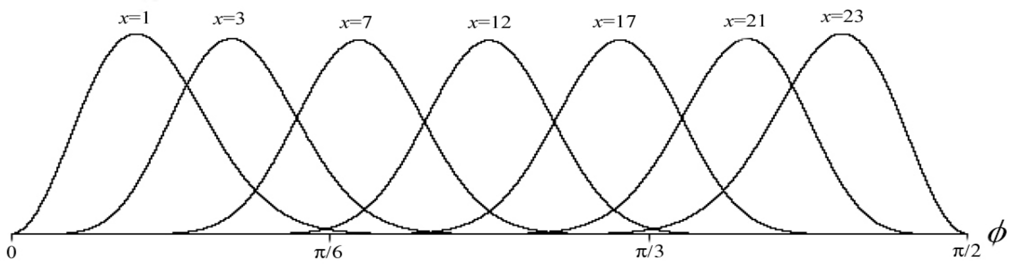

Figure 2. Standardized likelihood curves for $\phi = sin^{-1}\sqrt{\theta}$.

Notice that, although the superior profile of standardized likelihood for ϕ is nearly constant, the corresponding curves are rather far from being data translated. That suggests that the right criterion for the assumption of a uniform prior is based on the profile of likelihood rather than 'data translated likelihood'.

Given the general rule, defined by (8), we could build approximate prior distributions, by assuming each time the prior for θ:

$$g_r(\theta) \propto \sup_t g_{r-1}(\theta)\ell(\theta \mid t), \tag{12}$$

where $r = 1, 2, ..., s$, s is fixed in such a way that

$$g_s(\theta) \approx g_{s-1}(\theta), \tag{13}$$

and

$$g_o(\theta) = \frac{\ell'(\theta \mid t)}{\ell(\theta \mid t)}. \tag{14}$$

Of course, the iteration could go up to the asymptotic prior distribution, if it exists.

We note in passing that in the above-mentioned procedure the prior for $r = 1$ may approximate very well the prior for $r = 2$. That is, the approximation may be rather good already for $r = 1$.

A further good approximation for the prior concerning the parameter θ could be provided by the superior profile of the sampling distribution of the t given θ, with reference to all possible determinations of t from the intended design. detail, in a future paper.

The profile criterion could be extended to the case of two or more parameters. For instance, we can consider the choice of the prior with reference to a random sampling, in samples of size n, from a Normal distribution $N(\mu, \sigma^2)$, where μ and σ are both unknown. The likelihood of (μ, σ) is

$$\ell(\mu, \sigma \mid \mathbf{x}) \propto \sigma^{-n} \exp\Big\{ -\frac{n}{2\sigma^2}\Big[s^2 + (m-\mu)^2\Big]\Big\}, \tag{15}$$

where m and s are the maximum likelihood estimators about μ and σ, respectively. According to Box and Tiao (1973, p. 49), this likelihood is data translated (and the design impartial) in terms of $(\mu, \log\sigma)$. Then, following the profile criterion, the density prior for (μ, σ) is proportional to the Jacobian of the transformation from $(\mu, \log\sigma)$ to (μ, σ). That is, in accordance with the modified Jeffreys-rule, $p(\mu, \sigma) \propto \sigma^{-1}$.

Jeffreys has made fundamental contributions to statistics in order to obtain analytical prior distributions. Anyway, we believe that the basic idea for choice and evaluation of an objective prior is profile criterion.

In the end, it is clear that the thesis, supported by Howson and Urbach (1996, p. 429), according to which no theory can adequately represent 'ignorance' between alternatives, has to be revised in the light of the new principle of indifference.

4 An illustration of the profile criterion

4.1 Profile criterion prior for Binomial sampling

For a Binomial sample of size n with x successes, the standardized likelihood $\ell'(\theta \,|\, x)$ has been given in (10). In order to determine the profile criterion density defined by $\pi(\theta) \propto \sup_x \ell'(\theta \,|\, x)$, let us consider the ratio

$$R(y) = \frac{\ell'(\theta \,|\, x+1)}{\ell'(\theta \,|\, x)} = \frac{n-x}{x+1}\,\frac{\theta}{1-\theta}. \tag{16}$$

For a given x, this ratio is equal to one for $\theta = (x+1)/(n+1)$, is greater than one for $\theta < (x+1)/(n+1)$ and is smaller than one for $\theta > (x+1)/(n+1)$. Consequently we get

$$\pi(\theta) \propto \ell'\Big(\theta \,|\, X(\theta, n)\Big), \tag{17}$$

where

$$X(\theta, n) = j \quad \text{for} \quad \frac{j}{n+1} \leq \theta \ \leq \frac{j+1}{n+1} \quad (0 \leq j \leq n). \tag{18}$$

For θ in each interval $[j/(n{+}1), (j{+}1)/(n{+}1)]$ $(0 \leq j \leq n)$, $\pi(\theta)$ is proportional to the density of the Beta distribution $\beta(j{+}1, n{-}j{+}1)$. The fact that $R(j) = 1$ when $\theta = (j+1)/(n+1)$ ensures the continuity of the density. The constant of normalization $1/\int_0^1 \pi(\theta) d(\theta)$ can be easily computed from incomplete Beta functions. Moreover, recurrence relations can be derived to get more efficient computer algorithms. The prior for $n = 5$ is plotted in Figure 3 and can be compared with the Jeffreys prior. The constant of normalization is 1/3.018. It can be seen in Figure 3 that the transformed profile criterion prior for $\phi = sin^{-1}\sqrt{\theta}$ is approximately uniform on the interval [0.17,1.40], that is for θ between about 0.03 and 0.97.

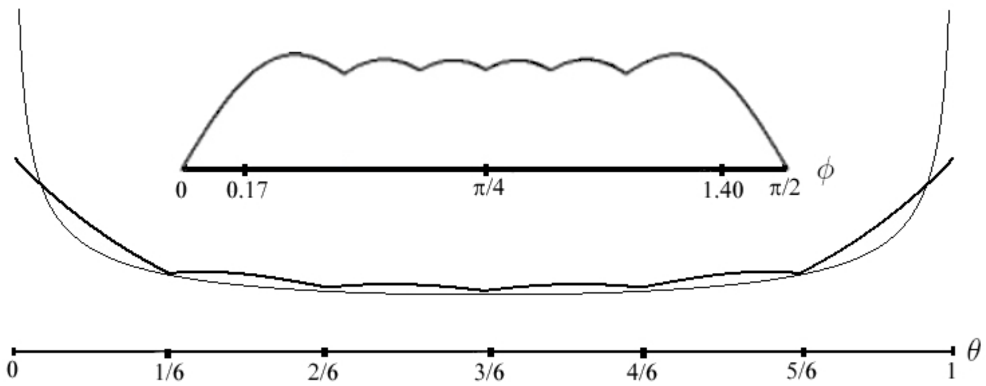

Figure 3. Profile criterion prior (thick line) and Jeffreys prior $\beta(\frac{1}{2},\frac{1}{2})$ (thin line) for a Binomial sample of size $n = 5$. The top curve is the transformed profile criterion prior for $\phi = sin^{-1}\sqrt{\theta}$.

4.2 Numerical applications

For a Binomial sample of size n, it follows that the corresponding posterior density $\pi(\theta \,|\, x)$ within each interval defined by j $(0 \leq j \leq n)$ is proportional to the density of the Beta distribution $\beta(x + j + 1, 2n - x - j + 1)$, with a coefficient of proportionality that depends on j and is equal to

$$\frac{B(x + j + 1, 2n - x - j + 1)}{B(j + 1, n - j + 1)}. \tag{19}$$

Here again the constant of normalization can be easily computed from incomplete Beta functions and recurrence relations can be derived, assuring the feasibility of the procedure. For instance, for $n = 5$ the posterior associated with $x = 2$ and $x = 0$ are plotted in Figure 4.

For $x = 2$ the posterior is very closed to the posterior Beta distribution $\beta(2.5, 3.5)$ associated with the Jeffreys prior. These distributions have respective medians 0.4070 and 0.4068. We get for instance the respective 95% equal tails credible intervals [0.0930,0.7894] and [0.0944,0.7906].

For the extreme case $x = 0$ the posterior introduces a correction to the Jeffreys posterior $\beta(0.5, 5.5)$. These two distributions have respective medians 0.070 and 0.042. We get for instance the respective 95% equal tails credible intervals [0.0025,0.4046] and [0.00009,0.3794].

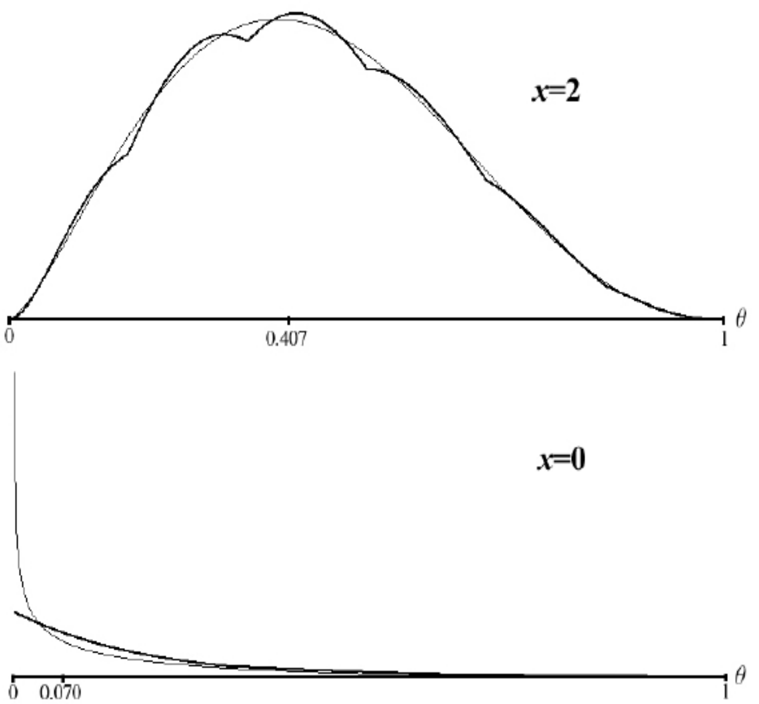

Figure 4. Posterior distributions associated with the profile criterion (thick line) and Jeffreys priors (thin line) for a Binomial sample of size $n = 5$ with $x = 2$ (top figure) and $x = 0$ successes (bottom figure).

4.3 Frequentist properties

One of the most common approaches to the evaluation of an objective (or neutral) prior distribution – at first developed by Welch and Peers (1963) – is to see whether it yields posterior credible sets that have good frequentist coverage properties. Actually, the Jeffreys credible interval has remarkable frequentist properties. Its coverage probability is very close to the nominal level, even for small-size samples, and it can be favorably compared to most frequentist intervals (Brown, Cai and DasGupta, 2001; Cai, 2005). So, Cai (2005) concluded: "The results show that the Jeffreys and second-order corrected intervals provide significant improvements over both the Wald and score intervals. These two alternative intervals nearly completely eliminate the systematic bias in the coverage probability. The one-sided Jeffreys and second-order corrected intervals can be resolutely recommended." (p.81).

It can be shown that the correction introduced by the profile criterion prior provides again better frequentist properties than the Jeffreys interval. Let us consider for illustration the same example as Cai (2005, pp. 69-70): the coverage probability of the 99% upper limit confidence interval for a binomial proportion with $n = 30$. Figure 5 plots the coverage probabilities of the profile criterion interval as a function of θ. In order to compare the performance of the intervals, let us introduce the following "optimal" coverage probabilities. For any value θ, let us define P^+, the smaller possible coverage probabilities superior to 99%, and P_-, the larger possible coverage probabilities inferior to 99%. P^+ and P_- are the coverage probabilities of the two 99% Bayesian posterior limits respectively associated with the priors $\beta(1,0)$ and $\beta(0,1)$ (Lecoutre, 2008, p. 790). Note that the first limit is the 99% upper limit of the Wald interval. For any θ, an optimal coverage probability must be either P^+ or P_-. We will conventionally consider as optimal the closer value to 99%. For the Jeffreys and profile criterion intervals, Figure plots the differences between the coverage probability and the optimal coverage probability, showing that the profile criterion leads to superior performance.

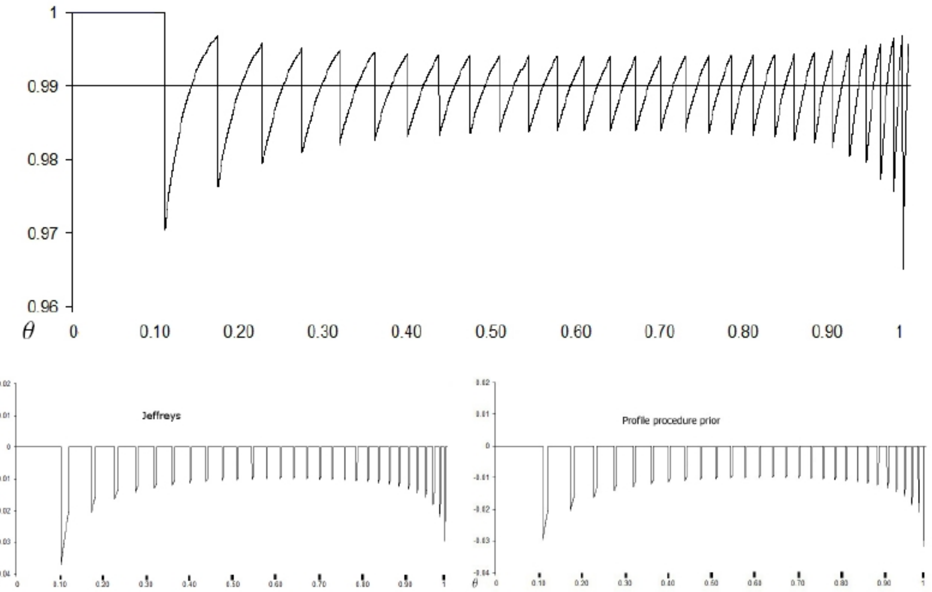

Figure 5. Coverage probability of the 99% upper limit (profile criterion prior) for a binomial proportion with $n = 30$ (top curve) and differences between the coverage probability (Jeffreys and profile criterion intervals) and the optimal coverage probability.

Note that, even for very extreme proportions, the coverage probability remains exceptionally good. So, for all values θ form 0.001 to 0.999 by step of 0.001, the "less good" coverage probabilities are respectively 0.9702 ($\theta = 0.995$) for the Jeffreys prior and 0.9417 ($\theta = 0.998$) for the "profile criterion". By contrast, for the second order corrected interval, we can get the unacceptable coverage probabilities 0.0825, 0.0566 and 0.0291 for $\theta = 0.997$, 0.998 and 0.999.

5 Conclusion

The new principle of indifference allows us to be consistent with probability axioms besides achieving the long sought-after objective of science: an induction that can be analytically controlled in all its constituent elements. In this way, we actually answer Hume's challenge, and, in the Carnap's words, we can take the degree of confirmation as the basic concept for all of inductive statistics.

This is not to deny the importance of the subjective theory of probability. In effect, the ideal of total objectivity is unattainable. Yet, the result of the inference can be notified in objective (or analytical) form, not only when prior knowledge is held to be irrelevant, but also in the other cases, making the subjective component of information (that determined the induction) explicit.

Quoting Fisher's words (1955), "we have the duty of formulating, of summarizing, and of communicating our conclusions, in intelligible form, in recognition of the right of other free minds to utilize them in making their own decisions".

Similarly, according to Carnap (1962), inductive logic alone does not and cannot determine the best hypothesis on a given evidence, if the best hypothesis means that which good scientists would prefer. It tell them to what degree the hypothesis considered is supported by the observations.

In other words, the most important task of probability calculus in statistics is to provide objective measures of evidence produced by experiments or other sample surveys, with an inference entirely probabilistic.

Anyway, much work still remains to be done in order to implement procedures and build objective prior distributions. Later, it should be opportune to review all the methods of inference, so concluding the induction with a probability distribution, opportunely analyzed for a description (possible exhaustive) of the inductive process.

In conclusion, a necessary and sufficient condition for consistency of scientific inference is not only the agreement with the rules of probability calculus, but also the *analytical solution* that the new principle of indifference is able to give the problem of statistical induction. In passing, this was the essential requisite of Carnap's work, or, rather, his own goal.

References

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

[2] Brown, L. D., Cai, T., and DasGupta, A. (2001). Interval estimation for a binomial proportion (with discussion). *Statistical Science*, **16**, 101–133.

[3] Bunouf, P., Lecoutre, B. (2006). Bayesian priors in sequential binomial design. *Comptes Rendus de L'Académie des Sciences de Paris, Série I*, **343**, 339–344.

[4] Bunouf P., Lecoutre B. (2008). On Bayesian estimators in multistage binomial designs. *Journal of Statistical Planning and inference*, **138**, 3915–3926.

[5] Bunouf P., Lecoutre B. (2010). An objective Bayesian approach to multistage hypothesis testing. *Sequential Analysis*, **29**, 88–101.

[6] Cai, T. (2005). One-sided confidence intervals in discrete distributions. *Journal of Statistical Planning and Inference*, **131**, 63–88.

[7] Carnap, R. (1962). *Logical Foundation of Probability.* 2^{nd} ed. Chicago: University Press.

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

[9] de Cristofaro, R. (2004). On the foundations of likelihood principle. *Journal of Statistical Planning and Inference*, **126**, 401–411.

[10] de Cristofaro, R. (2008). A new formulation of the principle of indifference. *Synthese*, **163**, 329–339.

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

[12] Fisher, R. A. (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society B*, **17**, 69–78.

[13] Howson, C. and Urbach, P. (1996). *Scientific Reasoning: The Bayesian Approach*, 2^{nd} edition. Chicago: Open Court Publishing Company.

[14] Lecoutre B. (2008). Bayesian methods for experimental data analysis. *Handbook of statistics: Epidemiology and Medical Statistics* (Vol 27), Amsterdam: Elsevier, 775–812.

[15] Lindley, D. V. (2000). The philosophy of statistics. *The Statistician*, **49**, 293–337.

[16] Severini, T. A. (2000). *Likelihood Methods in Statistics*, Oxford Statistical Science Series, 22. Oxford: University Press.

[17] Welch, B. and Peers, H. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *Journal of the Royal Statistical Society B*, 25, 318–329.