

Natural Induction: An Objective Bayesian Approach

James O. Berger, José M. Bernardo and Dongchu Sun

Abstract. The statistical analysis of a sample taken from a finite population is a classic problem for which no generally accepted objective Bayesian results seem to exist. Bayesian solutions to this problem may be very sensitive to the choice of the prior, and there is no consensus as to the appropriate prior to use.

This paper uses new developments in reference prior theory to justify and generalize Perks (1947) ([15]) ‘rule of succession’ — determining the probability that a new element from a population will have a property, given that *all* n previous elements from a random sample possessed the property — and to propose a new objective Bayesian solution to the ‘law of natural induction’ problem — determining the probability that all elements in a finite population have the property, given that all previous elements had the property.

The prior used for the first problem is the reference prior for an underlying hypergeometric probability model, a prior first suggested by Jeffreys (1946) ([10]) and recently justified on the basis of an exchangeability argument in Berger, Bernardo and Sun (2009) ([4]). The reference prior in the second problem arises as a modification to this prior that results from declaring the quantity of interest to be whether or not all the elements in the finite population have the property under scrutiny.

Inducción en las Ciencias de la Naturaleza: Una Solución Bayesiana Objetiva

Resumen. El análisis estadístico de una muestra aleatoria extraída de una población finita es un problema clásico para el que no parece existir una solución bayesiana generalmente aceptada. Las soluciones bayesianas a este problema pueden ser muy sensibles a la elección de la distribución inicial, y no existe consenso sobre la distribución inicial que debería ser utilizada.

En este trabajo se hace uso de desarrollos recientes del análisis de referencia para justificar y generalizar la solución de Perks (1947) ([15]) a la ‘regla de sucesión’ — la probabilidad de que un nuevo elemento de la población tenga una propiedad si *todos* los elementos de una muestra aleatoria la tienen — y para proponer una nueva solución bayesiana objetiva a la ‘ley de inducción natural’, — la probabilidad de que todos los elementos de una población finita tengan una propiedad si todos los elementos de la muestra la tienen. La distribución inicial utilizada para el primer problema es la distribución de referencia para el modelo probabilístico hipergeométrico subyacente, una distribución inicial sugerida por Jeffreys (1946) ([10]) y recientemente justificada utilizando un argumento de intercambiabilidad en Berger, Bernardo and Sun (2009) ([4]). La distribución de referencia para el segundo problema se obtiene

Presentado por / Submitted by Francisco Javier Girón González-Torre.

Recibido / Received: 12 de febrero de 2009. Aceptado / Accepted: 4 de marzo de 2009.

Palabras clave / Keywords: Bayes factors, Binomial model, Finite parameter space, Finite population sampling, Jeffreys prior, Hypergeometric model, Laplace law of succession, Natural induction, Objective Bayes, Precise hypothesis testing, Objective prior, Reference analysis, Reference prior.

Mathematics Subject Classifications: Primary: 62A01, 62F15; Secondary: 62B10, 62P10, 62P30, 62P35.

© 2009 Real Academia de Ciencias, España.

como resultado de modificar la distribución anterior al declarar que el problema de interés es determinar si es o no es cierto que todos los elementos de una población finita tienen la propiedad objeto de estudio.

1 The Problem

The “rule of succession” and “law of natural induction” (see below for definitions) have been discussed for hundreds of years by scientists, philosophers, mathematicians and statisticians. Zabell (1989, 2005) ([18, 19]) gives introductions to much of the more quantitative side of this history, as well as providing numerous modern insights.

Our focus here is primarily technical: to produce the specific “rule” and “law” that arise from adopting the *reference prior* approach to objective Bayesian analysis, as this approach has proven itself to be quite successful in a wide variety of contexts (see Bernardo, 1979, 2005 ([5, 7]); Berger and Bernardo, 1992 ([2]); and Berger, Bernardo and Sun, 2009 ([3, 4]), for discussion).

Sampling from a finite population. The most common statistical framework in which these subjects are discussed is that of a finite population of size N , where the interest centers on R , the unknown number of elements from the population which share a certain property. For instance, the population may consist of a batch of N recently produced items, R of which satisfy the required specifications and may therefore be safely sold, or it may consist of a population of N individuals, R of which share some genetic characteristic. The elements which share the property under analysis will be called *conforming*, and the event that a particular element in the population is conforming will be denoted by $+$. Given the information provided by a random sample of size n (without replacement) from the population which has yielded r conforming items, interest usually centers in one of these problems:

- The proportion $\theta = R/N$ of conforming items in the population.
- The probability $\Pr(+ | r, n, N)$ that an element randomly selected among the remaining unobserved $N - n$ elements turns out to be conforming. The particular case $\Pr(+ | r = n, n, N)$, that the next observed item is conforming, given that the first n observed elements are conforming, is commonly referred to as the *rule of succession*.
- The probability $\Pr(\text{All } + | n, N)$ that *all* the elements in the population are conforming given that the first n observed elements are conforming. This is commonly referred to as the *law of natural induction*.

The probability model for the relevant sampling mechanism is clearly hypergeometric, so that

$$\Pr(r | n, R, N) = \text{Hy}(r | n, R, N) = \frac{\binom{R}{r} \binom{N-R}{n-r}}{\binom{N}{n}}, \quad (1)$$

where $r \in \{0, \dots, \min(R, n)\}$, $0 \leq n \leq N$, and $0 \leq R \leq N$.

Bayesian solutions to the problems described above require specification of a prior distribution $\Pr(R | N)$ over the unknown number R of conforming elements in the population. As will later become evident, these solutions are quite sensitive to the particular choice of the prior $\Pr(R | N)$. Note, however, that non-Bayesian solutions to these problems are very problematical, in part because of the discreteness of the problem and the fact that the interesting data outcome — all n observed elements are conforming — is extreme and, in part, because of structural difficulties with non-Bayesian approaches in *confirming* a precise hypothesis such as “all the elements in a population are conforming.”

Predictive and posterior distributions. If $\Pr(R | N)$, $R = 0, \dots, N$, defines a prior distribution for R , the posterior probability of R conforming elements in the population having observed r conforming elements within a random sample of size n is, by Bayes theorem,

$$\Pr(R | r, n, N) = \frac{\text{Hy}(r | n, R, N) \Pr(R | N)}{\Pr(r | n, N)}, \quad (2)$$

for $R \in \{r, \dots, N - n + r\}$, and zero otherwise, where

$$\Pr(r | n, N) = \sum_{R=r}^{N-n+r} \text{Hy}(r | n, R, N) \Pr(R | N) \quad (3)$$

is the predictive distribution for the number r of conforming elements in a random sample of size n . Since, given N , the discrete parameter $\theta = R/N$, for $\theta \in \{0, 1/N, \dots, 1\}$, is a one-to-one transformation of R , the corresponding posterior of θ is just $\pi(\theta | r, N, n) = \Pr(N\theta | r, N, n)$.

Rule of succession. By the total probability theorem, the probability that an element randomly selected among the remaining unobserved $N - n$ elements is conforming is

$$\Pr(+ | r, n, N) = \sum_{R=r}^{N-n+r} \frac{R - r}{N - n} \Pr(R | r, n, N). \quad (4)$$

In particular, the probability of the event E_n that something which has occurred n times and has not hitherto failed to occur (so that $r = n$) will occur again is

$$\Pr(E_n | N) = \Pr(+ | r = n, n, N) \quad (5)$$

which, for many commonly used priors, turns out to be independent of the population size N . This is commonly referred to as the rule of succession. As n increases, $\Pr(E_n | N)$ converges quickly to one for all N for commonly used priors. This agrees with the usual perception that, if an event has been observed for a relatively large uninterrupted number of times, it is very likely that it will be observed again in the next occasion.

Law of natural induction. The posterior probability that all the N elements in the population are conforming given that all the n elements in the sample are, is

$$\Pr(\text{All} + | n, N) = \Pr(R = N | r = n, n, N). \quad (6)$$

In typical applications, n will be moderate and N will be much larger than n . For many conventional priors, $\Pr(\text{All} + | n, N)$ would then be very small, and this clearly conflicts with the common perception from scientists that, as n increases, $\Pr(\text{All} + | n, N)$ should converge to one, whatever the (often very large) value of N might be. A formal objective Bayesian solution to this problem, typically known as the law of natural induction, is the main objective of this paper, and will require following a suggestion of Jeffreys.

2 Conventional Objective Bayesian Inference

Both Bayes (1763) [1] and Laplace (1774, 1812) [13, 14] utilized a constant prior for unknowns. Since R may take $N + 1$ different values, the constant prior is thus the uniform distribution

$$\pi_u(R | N) = \frac{1}{N + 1}, \quad R = 0, \dots, N. \quad (7)$$

We first review the analysis for this conventional objective prior, as given in Broad (1918) ([9]).

Predictive and posterior distributions. The corresponding predictive distribution $\pi_u(r | n, N)$ for the number r of conforming elements in a random sample of size n is

$$\pi_u(r | n, N) = \frac{1}{N + 1} \sum_{R=r}^{N-n+r} \text{Hy}(r | n, R, N) = \frac{1}{n + 1},$$

for $r = 0, \dots, n$, a uniform distribution over its $n + 1$ possible values, which is therefore independent of N . Substituting into (2), the corresponding posterior distribution for R is

$$\pi_u(R | r, n, N) = \frac{n+1}{N+1} \text{Hy}(r | n, R, N) = \frac{\binom{R}{r} \binom{N-R}{n-r}}{\binom{N+1}{n+1}}, \quad (8)$$

for $R \in \{r, \dots, N - n + r\}$, and zero otherwise. In particular, the posterior probability that all the N elements in the population are conforming, given that all the n elements in the sample are conforming, is

$$\pi_u(\text{All} + | n, N) = \pi_u(R = N | r = n, n, N) = \frac{n+1}{N+1}, \quad (9)$$

which is essentially the ratio of the sample size to the population size. Notice that, when n is much smaller than N as will often be the case, $\pi_u(\text{All} + | n, N)$ will be close to zero even for large values of the sample size n .

Law of succession. From (4), the probability that an element randomly selected from among the remaining unobserved $N - n$ elements is conforming given the uniform prior (7), so that $\Pr(R | r, n, N)$ is given in (8), reduces to

$$\pi_u(+ | r, n, N) = \sum_{R=r}^{N-n+r} \frac{R-r}{N-n} \frac{\binom{R}{r} \binom{N-R}{n-r}}{\binom{N+1}{n+1}} = \frac{r+1}{n+2}. \quad (10)$$

This is usually known as Laplace's rule of succession, although Laplace (1774) ([13]) did not consider the case of finite N and only derived (10) for its continuous binomial approximation without apparently realizing that this is also an exact expression for finite N , a result established by Broad (1918) ([9]). In particular, with the uniform prior (7), the probability of the event E_n that something which has occurred n times and has not hitherto failed to occur (so that $r = n$) will occur again is

$$\pi_u(E_n) = \pi_u(+ | r = n, n, N) = \frac{n+1}{n+2}, \quad (11)$$

which is independent of the population size N . As n increases, $\pi_u(E_n)$ quickly converges to one.

Notice the dramatically different behaviour of the seemingly related Equations (9) and (11). In typical applications, n will be moderate and N will be much larger than n ; if this is the case, $\pi_u(\text{All} + | n, N)$ will be close to zero, but $\pi_u(E_n)$ will be close to one. Thus, if an event has been observed for a relatively large number of uninterrupted times, and the uniform prior (7) is used for both problems, one obtains that it is very likely that it will be observed again in the next occasion, but quite unlikely that it will *always* be observed in the future.

3 Reference Analysis of the Hypergeometric Model

The conventional use of a uniform prior for discrete parameters can ignore the structure of the problem under consideration. For the hypergeometric model, where the values of R are actual numbers, not merely labels, there is arguably a clear structure. Indeed, for large N it is well known that the hypergeometric distribution is essentially equivalent to the binomial distribution with parameter $\theta = R/N$. The objective prior for R should thus be compatible with the appropriate objective prior for θ in a Binomial $\text{Bi}(r | n, \theta)$ model; this is commonly chosen to be the corresponding Jeffreys (and reference) prior, which is the (proper) distribution

$$\pi_r(\theta) = \text{Be}(\theta | \frac{1}{2}, \frac{1}{2}) = \frac{1}{\pi} \frac{1}{\sqrt{\theta(1-\theta)}}, \quad 0 < \theta < 1. \quad (12)$$

As N increases, the uniform prior (7) remains uniform, and this is clearly not compatible with (12).

Similar reasoning led Jeffreys (1946, 1961) ([10, 12]) to suppose that the R conforming items arose as a random sample from a binomial population with parameter p (which can be thought of as the limiting value of $\theta = R/N$ as $N \rightarrow \infty$), and then assign p the Jeffreys prior in (12). This hierarchical structure can be justified on the basis of exchangeability, as observed in Berger, Bernardo and Sun (2009) ([4]) (which also generalized the approach to other discrete distributions). The resulting induced reference prior for R is

$$\pi_r(R | N) = \frac{1}{\pi} \frac{\Gamma(R + \frac{1}{2}) \Gamma(N - R + \frac{1}{2})}{\Gamma(R + 1) \Gamma(N - R + 1)}, \quad R \in \{0, 1, \dots, N\}, \quad (13)$$

which may also be written as

$$\pi_r(R | N) = f(R) f(N - R), \quad R \in \{0, 1, \dots, N\}, \quad (14)$$

where

$$f(y) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(y + 1/2)}{\Gamma(y + 1)}, \quad y \geq 0. \quad (15)$$

As will later become evident, the positive, strictly decreasing function defined by (15) occurs very frequently in the derivations associated to the problems analyzed in this paper. Since $\Gamma(1/2) = \sqrt{\pi}$, $f(0) = 1$. Moreover, using Stirling's approximation to the Gamma functions, it is easily seen that, for large y ,

$$f(y) \approx \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{y}},$$

so that if R and $N - R$ are both large, one has

$$\pi_r(R | N) \approx \frac{1}{\pi} \frac{1}{\sqrt{R(N - R)}}.$$

For $N = 1$, the reference prior (14) is the uniform prior $\pi_r(R | N = 1) = \{1/2, 1/2\}$, for $R \in \{0, 1\}$, as one would certainly expect, but this is the *only* case where the reference prior for the hypergeometric is uniform. Indeed, using Stirling's approximation for the Gamma functions in (13) one gets, in terms of $\theta = R/N$,

$$\pi_r(\theta | N) \approx \frac{1}{N + \frac{2}{\pi}} \text{Be} \left(\frac{N\theta + \frac{1}{\pi}}{N + \frac{2}{\pi}} \mid \frac{1}{2}, \frac{1}{2} \right), \quad \theta = 0, 1/N, \dots, 1, \quad (16)$$

which is basically proportional to $\text{Be}(\theta | \frac{1}{2}, \frac{1}{2})$, and hence compatible with the reference prior for the continuous limiting model $\text{Bi}(r | n, p)$ with $p = \lim_{N \rightarrow \infty} R/N$.

Reference predictive and posterior distributions. Using (3), the reference prior predictive distribution of the number r of conforming items in a random sample of size n is

$$\begin{aligned} \Pr(r | n, N) &= \sum_{R=0}^N \text{Hy}(r | R, N, n) \pi_r(R | N) \\ &= \frac{1}{\pi} \frac{\Gamma(r + \frac{1}{2}) \Gamma(n - r + \frac{1}{2})}{\Gamma(r + 1) \Gamma(n - r + 1)} \\ &= f(r) f(n - r) = \pi_r(r | n), \end{aligned}$$

which is independent of the population size N . Notice that, as in the case of the uniform prior, the reference prior predictive distribution of r given n has precisely the same mathematical form as the reference prior of R given N , $\pi_r(R | N)$.

Furthermore, using (2) and the last result, the reference posterior distribution of R turns out to be

$$\begin{aligned} \pi_r(R | r, n, N) &= \frac{c(r, n, N) \Gamma(R + \frac{1}{2}) \Gamma(N - R + \frac{1}{2})}{\Gamma(R - r + 1) \Gamma(N - R - (n - r) + 1)}, \quad (17) \\ c(r, n, N) &= \frac{\Gamma(n + 1) \Gamma(N - n + 1)}{\Gamma(N + 1) \Gamma(r + \frac{1}{2}) \Gamma(n - r + \frac{1}{2})}. \end{aligned}$$

In particular, if $N = 2$ and $n = 1$, this yields

$$\pi_r(R | r = 0, n = 1, N = 2) = \{3/4, 1/4, 0\}, \quad R = 0, 1, 2,$$

which may be compared with the corresponding result $\{2/3, 1/3, 0\}$, obtained from a uniform prior.

Substituting $R = N$ and $r = n$ into (17) and simplifying, the reference posterior probability that all elements in the population are conforming, given that all elements in the sample are conforming, is

$$\pi_r(\text{All} + | n, N) = \frac{\Gamma(N + 1/2)}{\Gamma(N + 1)} \frac{\Gamma(n + 1)}{\Gamma(n + 1/2)} = \frac{f(N)}{f(n)} \approx \sqrt{\frac{n}{N}}. \quad (18)$$

Thus, $\pi_r(\text{All} + | n, N)$ is basically the square root of the ratio of the sample size to the population size, a considerable contrast to the result $(n + 1)/(N + 1)$ obtained in (9) for the uniform prior.

Rule of succession. From (4), the probability that an element randomly selected from among the remaining unobserved $N - n$ elements is conforming, for the reference prior (13), reduces to

$$\pi_r(+ | r, n, N) = \sum_{R=r}^{N-n+r} \frac{R - r}{N - n} \pi_r(R | r, n, N) = \frac{r + 1/2}{n + 1}, \quad (19)$$

which is independent of N . Equation (19) provides the reference rule of succession, which was first obtained by Perks (1947) ([15]) although, following Laplace, he only derived it for the limiting binomial approximation ($N = \infty$ case), apparently not realizing that it was also an exact expression for any finite population size N .

The reference rule of succession (19) may be compared with Laplace's $(r + 1)/(n + 2)$ of (10). In particular, the corresponding reference probability of the event E_n — that something which has occurred n times and has not hitherto failed to occur will occur — is, for any population size N ,

$$\pi_r(E_n) = \pi_r(+ | r = n, n, N) = \frac{n + 1/2}{n + 1}. \quad (20)$$

As one would require, for $n = 0$ (and hence with no initial information), both (11) and (20) yield $1/2$. For $n = 1$, Laplace yields $2/3$ while the corresponding reference probability is $3/4$. It is easily verified that, as n increases, the reference law of succession (20) has an appreciably faster convergence to one than Laplace's (11).

4 Natural Induction

In line with his early discussions on scientific enquiry (Wrinch and Jeffreys, 1921-23 ([17]), later expanded in Jeffreys, 1931 ([11]), Jeffreys (1961, p. 128) ([12]) disagreed with the result (9) — and would also have disagreed with (18) — arguing that, to justify natural induction, one should be able to demonstrate that a law is probably correct for all elements of a population of size N , given that it has proven to be correct in all of a very large number n of randomly chosen instances, even if n is appreciably smaller than N . In contrast, (9) and (18) can be quite small for large n if N is much larger than n . (Note that both (11) and (20) are near 1 for large n , but these probabilities refer to the event E_n that a further randomly chosen element will obey the stated law, not to the event that *all* elements in the population obey that law.)

To correct this problem, Jeffreys argued that the prior probability that all elements of the population have the property, $\Pr(R = N)$, must be some fixed value independent of N . He argued that this is reasonable, asserting that any clearly stated natural law has a positive prior probability of being correct, and he made several specific proposals (Jeffreys, 1961, Sec. 3.2 ([12])) for the choice of prior probability. The simplest choice is to let $\Pr(R = N) = 1/2$, and this is the choice arising from the reference analysis below. For a recent review of Jeffreys' ideas, see Robert, Chopin and Rousseau (2009) ([16]).

Reference analysis. A solution to the natural induction problem that satisfies the scientific desiderata described by Jeffreys may be obtained from a standard use of reference analysis, if the parameter of interest is chosen to be whether or not $R = N$, rather than the actual value of R . The result, described below, can also be phrased in terms of testing the hypothesis that $R = N$ versus the alternative $R \neq N$.

Lemma 1 Define the parameter of interest (in the reference prior analysis) to be

$$\phi = \begin{cases} \phi_0 & \text{if } R = N \text{ (All +),} \\ \phi_1 & \text{if } 0 \leq R < N. \end{cases}$$

Then the corresponding reference prior, $\pi_\phi(R | N)$, of the unknown parameter R is

$$\pi_\phi(R | N) = \begin{cases} \frac{1}{2} & \text{if } R = N \\ \frac{1}{2} \frac{f(R)f(N-R)}{1-f(N)} & \text{if } 0 \leq R < N, \end{cases}$$

where $f(y)$ is defined in (15).

PROOF. To have a representation of the unknown parameter R in terms of the quantity of interest ϕ and a nuisance parameter λ , define

$$\lambda = \begin{cases} \lambda_0 & \text{if } R = N \text{ (All +),} \\ R & \text{if } 0 \leq R < N. \end{cases}$$

The sampling distribution of the data r given $\phi = \phi_1$ is still the density

$$\Pr(r | \phi = \phi_1, \lambda \neq \lambda_0, n, N) = \text{Hy}(r | n, R, N),$$

but now R is restricted to the set $\{0, 1, \dots, N-1\}$. The reference prior of a model with a restricted parameter space is obtained as the restriction of the reference prior from the unrestricted model. Hence, using the fact that

$$\pi_r(N | N) = \frac{\Gamma(N+1/2)}{\sqrt{\pi} \Gamma(N+1)} = f(N),$$

the conditional reference prior of λ given $\phi = \phi_1$ is the renormalized version of (14)

$$\pi(\lambda | \phi = \phi_1, N) = \frac{f(R)f(N-R)}{1-f(N)}. \quad (21)$$

On the other hand, $\pi(\lambda = \lambda_0 | \phi = \phi_0, N) = 1$ since, given $\phi = \phi_0$, the nuisance parameter λ must be equal to λ_0 . Moreover, ϕ has only two possible values and, therefore, its marginal reference prior is simply $\pi(\phi = \phi_0) = \pi(\phi = \phi_1) = 1/2$. Hence, the joint reference prior of the unknown parameter R , when ϕ is the quantity of interest, is $\pi_\phi(R | N) = \pi(\lambda | \phi, N) \pi(\phi)$, which yields the conclusion. ■

Using this prior, the required reference posterior probability of the event $\{\text{All +}\}$ that all N elements in the population are conforming ($R = N$), given that all r elements in a random sample of size n are conforming ($r = n$), is found in the following result.

Theorem 1

$$\pi_\phi(\text{All +} | n, N) = \left(1 + \frac{f(n) - f(N)}{1 - f(N)}\right)^{-1}, \quad (22)$$

where $f(y)$ is defined in (15).

PROOF. Note that

$$\begin{aligned}\pi_\phi(\text{All} + | n, N) &= \pi_\phi(\phi = \phi_0 | r = n, N) \\ &= \frac{\frac{1}{2} \Pr(r = n | \phi = \phi_0, n, N)}{\frac{1}{2} \Pr(r = n | \phi = \phi_0, n, N) + \frac{1}{2} \Pr(r = n | \phi = \phi_1, n, N)} \\ &= (1 + \Pr(r = n | \phi = \phi_1, n, N))^{-1},\end{aligned}$$

since $\Pr(r = n | \phi = \phi_0, n, N)$ is obviously one. By the total probability theorem,

$$\Pr(n | \phi = \phi_1, n, N) = \sum_{R=n}^{N-1} \Pr(n | \phi = \phi_1, R, n, N) \pi_\phi(R | \phi = \phi_1, N)$$

and, using the fact that, for all N ,

$$\Pr(r = n | n, N) = \sum_{R=n}^N \text{Hy}(n | n, R, N) \pi_r(R | N, \mathcal{M}) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(n + \frac{1}{2})}{\Gamma(n + 1)} = f(n),$$

it is easily shown that

$$\Pr(r = n | R < N, n, N) = \frac{f(n) - f(N)}{1 - f(N)}.$$

The conclusion is immediate. ■

Hence, for large population sizes, $f(N) \approx 0$ and, as a consequence, the reference posterior probability, $\pi_\phi(R = N | r = n, N)$, that all N elements in the population are conforming, given that all elements in a sample of size n are conforming, is then essentially independent of N and given by

$$\pi_\phi(\text{All} + | n, N) \approx (1 + f(n))^{-1} = \left(1 + \frac{1}{\sqrt{\pi}} \frac{\Gamma(n + \frac{1}{2})}{\Gamma(n + 1)}\right)^{-1}, \quad (23)$$

which, for moderately large n , may further be approximated by

$$\pi_\phi(\text{All} + | n, N) \approx \frac{\sqrt{n}}{\pi^{-1/2} + \sqrt{n}}. \quad (24)$$

For instance, with $n = 100$ and $N = 1000$ the exact value of the required probability, given by (22), is $\pi_\phi(R = N | r = n, N) = 0.9623$, and the two approximations (23) and (24), respectively, yield 0.9467 and 0.9466.

Equation (22) may be compared with the result which one would obtain if a conventional uniform conditional prior

$$\pi_1(\lambda | \phi = \phi_1, N) = \pi_1(R | N - 1) = \frac{1}{N},$$

had been used instead of the structured conditional reference prior (21). It may be shown (Bernardo, 1985 ([6]), Bernardo and Smith, 1994, p. 322 ([8])) that this yields

$$\pi_1(\text{All} + | n, N) = \left(1 + \frac{1}{n+1} \left(1 - \frac{n}{N}\right)\right)^{-1}, \quad (25)$$

which is always larger than (22). For instance, with $n = 50$ and $N = 100$, the result in Theorem 1 yields $\pi_\phi(\text{All} + | n = 50, N = 100) = 0.9263$, while the use of (25) yields $\pi_1(\text{All} + | n = 50, N = 100) = 0.9903$, a much larger value. Thus the reference probability is considerably more conservative.

$N = \infty$ and hypothesis testing. As mentioned earlier, as $N \rightarrow \infty$, the hypergeometric $\text{Hy}(r | n, R, N)$ model converges to the binomial $\text{Bi}(r | n, p)$ model, with $p = \lim_{N \rightarrow \infty} R/N$. In this infinite population

setting, the model that all elements in the population have the specified property can be stated in the language of hypothesis testing as $H_0 \equiv \{p = 1\}$. The natural induction problem is thus formally that of finding the posterior probability of H_0 when all n observations have the property (i.e., are successes in the binomial model).

The reference prior assigns probability $1/2$ to both $H_0 \equiv \{p = 1\}$ and $H_1 \equiv \{p < 1\}$ and, conditional on H_1 , assigns p the $\text{Be}(p | \frac{1}{2}, \frac{1}{2})$ reference prior. A straightforward computation then yields that the posterior probability of H_0 , given that the first n observations all have the specified property, is equal to

$$\Pr(p = 1 | n) = \left(1 + \frac{1}{\sqrt{\pi}} \frac{\Gamma(n + \frac{1}{2})}{\Gamma(n + 1)}\right)^{-1} = (1 + f(n))^{-1} \quad (26)$$

which, curiously, is the same expression as the approximation to the reference posterior probability $\pi_\phi(\text{All} + | n, N)$ obtained for large N in (23).

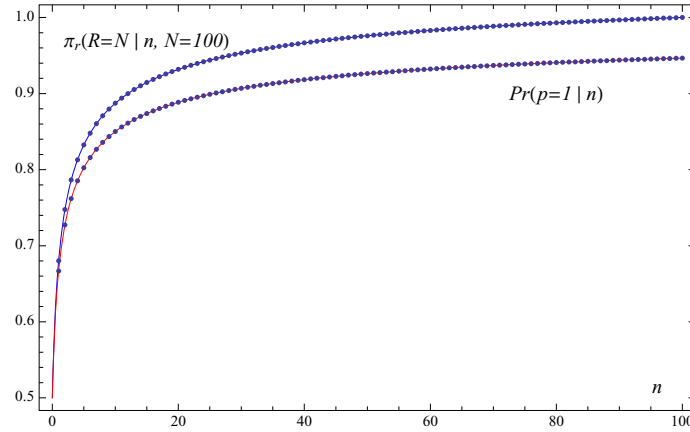


Figure 1. Reference posterior probabilities $\pi_\phi(\text{All} + | n, N = 100)$ that all the elements from a finite population of size $N = 100$ are conforming, given that all the elements from a random sample of size n are conforming, for $n = 0, 1, \dots, 100$, (above) and its Bayes factor continuous ($N = \infty$) approximation (below).

Figure 1 shows the exact values of (22) for $N = 100$ and $n = 0, 1, \dots, 100$, along with the binomial limiting case. Note that it is important to take the population size, N , into account, except for very large population sizes. For example, if $r = 50$ and $N = 100$, the exact value of the required reference posterior probability that all elements in the population are conforming is $\pi_\phi(\text{All} + | n, N) = 0.9617$, but the continuous approximation gives only $\Pr(p = 1 | n) = 0.9263$.

5 Conclusions and an Example

The proposed solutions, based on reference prior theory, to the two originally posed problems can be summarized as follows.

The Reference Rule of Succession. In a population of N elements, from which n have been randomly sampled and been found to be conforming, the reference posterior probability that a new element, randomly selected from among the remaining unobserved $N - n$ elements, turns out to be conforming is

$$\pi_r(+ | r = n, n, N) = \frac{n + 1/2}{n + 1}, \quad (27)$$

which is independent of N and may be compared with Laplace's $(n + 1)/(n + 2)$.

The Reference Law of Natural Induction. If it is assumed that the parameter of interest is whether or not *all* elements of a population are conforming (i.e. that the property is a law of nature), then the reference posterior probability of this law is

$$\pi_{\phi}(\text{All} + | n, N) = \left(1 + \frac{\frac{1}{\sqrt{\pi}} \frac{\Gamma(n+1/2)}{\Gamma(n+1)} - \frac{1}{\sqrt{\pi}} \frac{\Gamma(N+1/2)}{\Gamma(N+1)}}{1 - \frac{1}{\sqrt{\pi}} \frac{\Gamma(N+1/2)}{\Gamma(N+1)}} \right)^{-1}. \quad (28)$$

For very large N (or infinite populations), this is essentially identical to (26).

Example. Many years ago, when visiting the Charles Darwin research station of the Galápagos islands, one of us (Bernardo) had a question posed by a zoologist. The zoologist had observed, and marked, 55 galápagos (tortoises) in one small island, all of which presented a particular shell modification, possibly related to their adaptation to the island vegetation. The zoologist asked for the probability that all galápagos in the island had that modification. He added that his sample was roughly random, and that he would estimate the total galápagos population in that island to be between 150 and 250 individuals.

At the time, Bernardo quoted the solution (25) based on a conditional uniform prior, which yielded the range $[0.986, 0.989]$ (corresponding to the range of N). The reference probabilities (28) give the smaller values

$$\pi_{\phi}(\text{All} + | n = 55, N \in [150, 250]) \in [0.960, 0.970].$$

Note that these numbers are, nevertheless, appreciably higher than the limiting binomial approximation $\Pr(p = 1 | n = 55) = 0.929$, showing the importance of incorporating N into the analysis.

Finally, using (27), the zoologist could have been told that, if he went back to the island, he would have a reference posterior probability

$$\pi_r(+ | r = n, n = 55, N) = 0.991$$

that the first found unmarked galápagos also presented a modified shell.

Acknowledgement. J. O. Berger was supported by grant DMS-0635449 of the National Science Foundation, USA, J. M. Bernardo was supported by grant MTM2006-07801 of the Ministerio de Educación y Ciencia, Spain, and D. Sun was supported by grant SES-0720229 of the National Science Foundation, USA, and grants R01-MH071418 and R01-CA109675 of the National Institute of Health, USA.

References

- [1] BAYES, T., (1763). An essay towards solving a problem in the doctrine of chances. Published posthumously in *Phil. Trans. Roy. Soc. London*, **53**, 370–418 and **54**, 296–325.
- [2] BERGER, J. O. AND BERNARDO, J. M. (1992). On the development of reference priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.), Oxford: University Press, 35–60 (with discussion).
- [3] BERGER, J., BERNARDO, J. M. AND SUN, D. (2009). The formal definition of reference priors. *Ann. Statist.*, **37**, 905–938.
- [4] BERGER, J., BERNARDO, J. M. AND SUN, D. (2009). Reference priors for discrete parameter spaces. *Submitted*.
- [5] BERNARDO, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. B*, **41**, 113–147 (with discussion). Reprinted in *Bayesian Inference*, **1** (G. C. Tiao and N. G. Polson, eds.) Oxford: Edward Elgar, 229–263.
- [6] BERNARDO, J. M. (1985). On a famous problem of induction. *Trab. Estadist.*, **36**, 24–30.

- [7] BERNARDO, J. M. (2005). Reference analysis. *Handbook of Statistics*, **25**, (D. K. Dey and C. R. Rao, eds.) Amsterdam: Elsevier, 17–90.
- [8] BERNARDO, J. M. AND SMITH, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley.
- [9] BROAD, C. D. (1918). On the relation between induction and probability. *Mind*, **27**, 389–404.
- [10] JEFFREYS, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Royal Soc. London A*, **186**, 453–461.
- [11] JEFFREYS, H. (1931). *Scientific Inference*. Cambridge: University Press.
- [12] JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Oxford: University Press.
- [13] LAPLACE, P. S. (1774). Mémoire sur la probabilité des causes par les événements. *Mém. Acad. Sci. Paris*, **6**, 621–656. Reprinted in *Oeuvres Complètes* **8**, 27–68. Paris: Gauthier-Villars, 1891.
- [14] LAPLACE, P. S. (1812). *Théorie Analytique des Probabilités*. Paris: Courcier. Reprinted in *Oeuvres Complètes de Laplace*, **7**, 1878–1912. Paris: Gauthier-Villars.
- [15] PERKS, W. (1947). Some observations on inverse probability including a new indifference rule. *J. Inst. Actuaries* **73**, 285–334 (with discussion).
- [16] ROBERT, C. P., CHOPIN N. AND ROUSSEAU, J. (2009). Harold Jeffreys' Theory of Probability revisited. *Statist. Science*, **23**, (to appear, with discussion).
- [17] WRINCH, D. AND JEFFREYS, H. (1921-23). On certain fundamental principles of scientific inquiry. *Philosophical Magazine* **42**, 369–390 and **45**, 368–374.
- [18] ZABELL, S. L. (1989). The rule of succession. *Erkenntnis*, **31**, 283–321.
- [19] ZABELL, S. L. (2005). *Symmetry and Its Discontents*. Cambridge: University Press.

James O. Berger

Dept. Statistical Science,
Duke University,
USA

berger@stat.duke.edu

José M. Bernardo

Dept. Estadística e I.O.,
Universitat de València,
Spain

jose.m.bernardo@uv.es

Dongchu Sun

Dept. Statistics,
University of Missouri-Columbia,
USA

sund@missouri.edu