WILEY
Publishers Since 1807

111 RIVER STREET, HOBOKEN, NJ 07030

Please follow these instructions to avoid delay of publication.

☐ **READ PROOFS CAREFULLY**

- This will be your only chance to review these proofs. **Please note that once your corrected article is posted online, it is considered legally published, and cannot be removed from the Web site for further corrections.**
- Please note that the volume and page numbers shown on the proofs are for position only.

☐ **ANSWER ALL QUERIES ON PROOFS** (Queries for you to answer are attached as the last page of your proof.)

- Mark all corrections directly on the proofs. Note that excessive author alterations may ultimately result in delay of publication and extra costs may be charged to you.

☐ **CHECK FIGURES AND TABLES CAREFULLY** (Color figure proofs will be sent under separate cover.)

- Check size, numbering, and orientation of figures.
- All images in the PDF are downsampled (reduced to lower resolution and file size) to facilitate Internet delivery. These images will appear at higher resolution and sharpness in the printed article.
- Review figure legends to ensure that they are complete.
- Check all tables. Review layout, title, and footnotes.

☐ **COMPLETE REPRINT ORDER FORM**

- Fill out the attached reprint order form. It is important to return the form even if you are not ordering reprints. You may, if you wish, pay for the reprints with a credit card. Reprints will be mailed only after your article appears in print. This is the most opportune time to order reprints. If you wait until after your article comes off press, the reprints will be considerably more expensive.

RETURN ☐**PROOFS**
☐**REPRINT ORDER FORM**

PLEASE RETURN PROOFS WITHIN 48 HOURS OF RECEIPT TO:

QUESTIONS?

Melissa Donnelly, Production Editor
Phone: 201-748-6438
FAX: 201-748-6052
E-mail: mdonnelly@wiley.com
Refer to journal acronym and article production number
(i.e., WICS for WIRES: Computational Statistics).

WILEY
Publishers Since 1807

REPRINT BILLING DEPARTMENT • 111 RIVER STREET • HOBOKEN, NJ 07030
PHONE: (201) 748-8789; FAX: (201) 748-6326
E-MAIL: reprints@wiley.com

PREPUBLICATION REPRINT ORDER FORM

Please complete this form if you wish to order reprints prior to publication. Please return this form with your proofs, even if you do not wish to order reprints. Your reprints will be shipped approximately 4 to 6 weeks after publication. Reprints ordered after printing are substantially more expensive.

JOURNAL: ***WIRES: Computational Statistics*** ______ VOLUME ______. ISSUE ______
TITLE OF MANUSCRIPT ______________________________.
MS. NO. ______. NO. OF PAGES ______. AUTHOR (S) ______________________.

REPRINTS 8 ¼ X 11

No. of Pages	100 Reprints	200 Reprints	300 Reprints	400 Reprints	500 Reprints
	$	$	$	$	$
1-4	336	501	694	890	1,052
5-8	469	703	987	1,251	1,477
9-12	594	923	1,234	1,565	1,850
13-16	714	1,156	1,527	1,901	2,273
17-20	794	1,340	1,775	2,212	2,648
21-24	911	1,529	2,031	2,536	3,037
25-28	1,004	1,707	2,267	2,828	3,388
29-32	1,108	1,894	2,515	3,135	3,755
33-36	1,219	2,092	2,773	3,456	4,143
37-40	1,329	2,290	3,033	3,776	4,528

** REPRINTS ARE ONLY AVAILABLE IN LOTS OF 100. IF YOU WISH TO ORDER MORE THAN 500 REPRINTS, PLEASE CONTACT OUR REPRINTS DEPARTMENT AT (201) 748-8789 FOR A PRICE QUOTE.

◈ Please send me ______. reprints of the above article at ……………	$ ______________.
Please add appropriate State and Local Tax {Tax Exempt No.____________. }	$ ______________.
Please add 5% Postage and Handling ………………………………………………	$ ______________.
TOTAL AMOUNT OF ORDER ** ………………………………………………	$ ______________.

*** International orders must be paid in U.S. currency and drawn on a U.S. bank*

Please check one: ◇ Check Enclosed ◇ Bill Me ◇ Credit Card
If credit card order, charge to: ◇ American Express ◇ Visa ◇ MasterCard ◇ Discover
Credit Card No. ______________. Signature ______________. Exp. Date __________.

Bill To:

Name ______________________.
Address ______________________.
______________________.
______________________.
______________________.
Purchase Order No.______________________.

Ship To:

Name ______________________.
Address ______________________.
______________________.
______________________.
______________________.
Phone ______________. Fax ______________.
Internet ______________________.

WILEY
Publishers Since 1807

WIRES: Computational Statistics (WICS)

To: Melissa Donnelly
Production Editor
Phone: 201-748-6438
Fax: 201-748-6052
From:
Date:
Pages including this cover page:

re:

c

Overview

Bayesian inference: an approach to statistical inference

D.A.S. Fraser*

an analogy involving/

and a statistical/

The original Bayes used an invariant prior ~~with a location~~ model and argued that the resulting combination of prior with likelihood provided a probability description of an unknown parameter value in an application; the particular combination in ~~the~~ context with invariance can currently be called a confidence distribution and is subject to some restrictions when used to construct confidence intervals and regions. The procedure of using a prior with likelihood has now, however, been widely generalized with invariance being ~~replaced by~~ less restrictive criteria such as non-informative, reference, and ~~others~~. Other generalizations are to allow the prior to represent various forms of background information that is ~~readily~~ available or elicited from those familiar with the statistical context; these can reasonably be called subjective priors. Still further generalizations address an anomaly where marginalization ~~for~~ a vector parameter gives results that contradict the term probability; these are Dawid, Stone, Zidek marginalization paradoxes; various priors for this are called targeted priors. ~~The~~ case where the prior describes ~~the~~ random source ~~of~~ the parameter value is just probability analysis but is frequently treated as a Bayes procedure. We survey the argument in support of probability characteristics and outline various generalizations of the original Bayes's proposal~~ity~~. © 2010 John Wiley & Sons, Inc. *WIREs Comp Stat* 2010 2 000–0000

tr/

∂/s/

extended to/

more/

∂/

with/

A special/

however/

a/for/

∂//

AQ2

can be applied to/

Bayesian inference is a form of statistical inference that has evolved from a seminal paper by Bayes.[1] In that paper, ~~Bayes considered~~ a special location type of statistical model: a statistical model $f(y;\theta)$ records possible density functions for a response variable y with uncertainty given by a parameter θ that manipulates the form of the density to present the perceived possibilities for the response distribution; and the particular location type of model ~~considered by Bayes had~~ the special form $f(y-\theta)$, which can be interpreted as a density for a variable z translated by an amount θ giving the response y as the translation $y=\theta+z$ of the initial z.

The proposal in/

has/

Bayes suggested that possible θ values could be viewed as coming from a prior density $\pi(\theta)$ that was constant in value, say $\pi(\theta)=c$, thus reflecting the translation invariance implicit in the location structure of the model. And then, if the prior distribution is taken to be descriptive of how the actual θ value in the application had been generated, we would obtain a joint distribution that described the pair (θ,y), and when observed data y^0 on the response y became available when standard conditional probability calculus would give the conditional distribution

t/

$$\pi(\theta|y^0)=c\pi(\theta)f(y^0;\theta) \tag{1}$$

where c is now the normalizing constant. This is called the posterior distribution for θ, and it would give a frequency description of possible values for the θ among instances (θ,y) where the observed y is equal to y^0, provided the prior is descriptive.

The Bayes approach was prominently endorsed by Laplace[2] and some other mathematicians and vigorously rejected by others, but it has had a profound influence on the development of statistical inference. Also, generalizations of the approach have become available for arbitrary statistical models $f(y;\theta)$ using priors $\pi(\theta)$ of various levels of appropriateness and relevance. And typically all are accompanied by some claim or assertion that the resulting posterior distribution $\pi(\theta|y^0)=c\pi(\theta)f(y^0;\theta)$ presents probabilities that in some way describe the unknown θ in the application.

---

*Correspondence to: dfraser@utstat.toronto.edu

AQ1 Department• of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3

DOI: 10.1002/wics.102

OK DF

THE INNOVATION

Bayes[1] proposal combined an observed statistical model $f^0(\theta) = f(y^0; \theta)$ coming from data y^0 with a weight function $\pi(\theta)$ that provided a calibration of parameter values. The use of the observed statistical model does appear as an extraordinarily relevant first step when a data value y^0 is obtained in the presence of a statistical model $f(y; \theta)$: just examine the model at the y^0-section through the model. And indeed Bayes' proposal made that model section the operating ingredient for formal statistical inference more than a century and a half before such section was formally conceptualized and given the name likelihood[3] and before the section was recommended[4] for direct plotting and viewing as an explicit presentation of information concerning the• parameter.

of

OK AQ3

The weight function $\pi(\theta)$ provides a calibration of the scaling of the parameter; and the translation invariance of the special model then leads to a constant prior density. And then if the prior density is viewed as being descriptive or having frequency properties, we would find that conditional probability calculations are in order. But, as proposed, the prior is just a mathematical object without reference to frequencies; accordingly something derived from it by probability calculus would also be just a mathematical object. Thus, a claim of consequent probability status is without corresponding basis. Indeed there are indications[3] that Bayes may have been hesitant to directly publish the proposal. Yet curiously the posterior distribution sort of works, sort of gives reasonable looking and behaving answers, and indeed is widely promoted.• But if it is not probability, then what properties underlie the frequent reasonable behavior of the Bayes posteriors?

location/

AQ4 ✓

ly found/

Consider the ~~original~~ case ~~as examined by~~ Bayes, but in current notation. Let $y = \theta + z$ where z has the centered distribution $g(z)$ of the model; for expository convenience to enable plotting we use the extreme value distribution $g(z) = e^{-z}\exp\{-e^{-z}\}$ which is simple and yet avoids having the special symmetries of the normal; see panel (a) in Figure 1. Then with data y^0, the calculation of the probability position of the data y^0 in the model with parameter value θ gives what can be viewed as the standard p-value function $p(\theta)$; it records the probability position of the data y^0 in the distribution centered at θ and is given by the observed value $F^0(\theta) = F(y^0 - \theta)$ of the distribution function. For the particular extreme value model, this is easily calculated and is given by

special/ and the /proposal/

$$p(\theta) = \exp\{-e^{-(y^0-\theta)}\} \quad (2)$$

which is plotted in Figure 2.

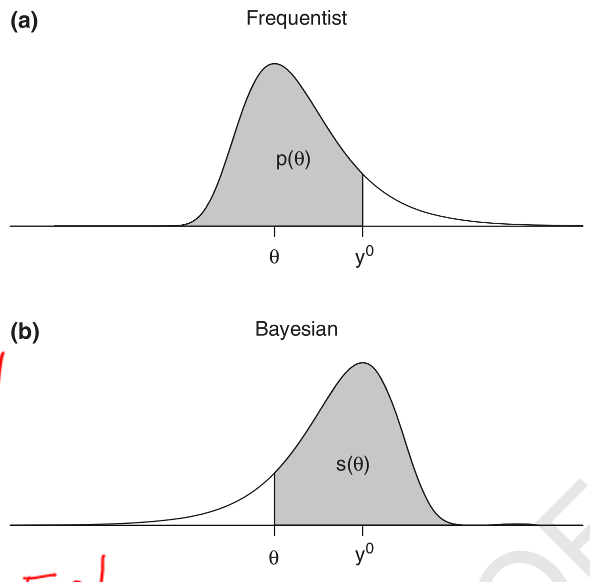

FIGURE 1 | The extreme value EV(θ,1) model: (a) the distribution of y given θ; (b) the posterior distribution of θ given y^0; the p-value $p(\theta)$ from panel (a) is equal to the survivor value $s(\theta)$ in panel (b).

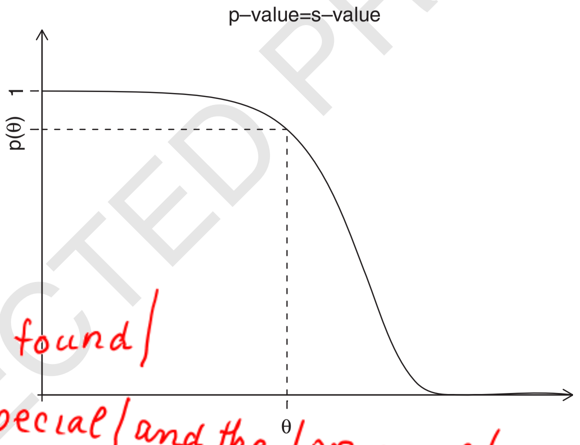

FIGURE 2 | •.

AQ5

If, however, we follow the Bayes proposal we would start with the likelihood function from the data y^0 and combine it with the indicated invariant Bayes prior $\pi(\theta) = c$; this in turn would give the Bayes posterior density: the result is just the normed likelihood function plotted in panel (b) of Figure 1; this is a reversed or mirrored extreme value shape, centered at the data point y^0. If we then calculate the right-tailed distribution function or survivor function from this posterior density, we obtain directly the same function, $s(\theta) = p(\theta)$, and thus would have the same reported information concerning the parameter θ.

The preceding information can also be put in another form. If we want the β-level lower confidence

✓

OK

AQ7

bound or the β-level posterior bound, we would calculate the θ value that makes $p(\theta)$ or $s(\theta)$ equal to β; we designate such quantile bound as $\hat{\theta}_\beta$. For the example, we then obtain

$$\hat{\theta}_\beta = y^0 + \log(-\log\beta) \qquad (3)$$

thus the posterior β lower bound for Bayes' original location model context is just the Fisher[5] lower confidence solution as clarified by Neyman.[6] And thus Bayes produced the Fisher--Neyman confidence result by a different route but by, unfortunately, a route that seems restricted to the location model case. We view this as the intrinsic explanation for the reasonable looking and behaving results from the Bayes procedure: the procedure gives confidence when the model has the nice linearity exemplified by the location models.

In summary, Bayes introduced the likelihood function as an essential ingredient for inference without giving it its subsequent name; and introduced confidence without giving it its subsequent name and used a derivation route which did not generalize beyond the simple location models. Clearly, these were major innovations even if not quite in the fully developed modern form now present more than two centuries later.• Perhaps the Bayes procedure can be viewed as an innovative method to obtain simple approximations for confidence.

AQ6

OK

PROBABILITY: BELIEF OR SUBSTANCE

The Bayes proposal focused on a prior probability density $\pi(\theta)$ to describe possible values for an unknown parameter value θ. In what way does it describe such values for a parameter? Is it asserting that the prior is objectively descriptive, and thus describing how the unknown value was randomly produced? In other words, are the probabilities in the prior real, objective and not just named or just mathematical or just for convenience? If the prior is objectively descriptive then the use of the conditional probability calculus would be appropriate, and the filtered pattern $\pi(\theta|y^0)$ would describe the possible θ value among cases (θ, y) where the variable y takes the observed value y^0, to some reasonable approximately. In other words, if $\pi(\theta)$ is objectively descriptive then $\pi(\theta|y^0)$ is objectively descriptive.

But if $\pi(\theta)$ is just a mathematical object representing judgments or personal views or opinions or other non-objective properties then correspondingly the posterior $\pi(\theta|y^0)$ is just a mathematical object at most representing judgments or personal views or opinions: the conditional calculus does not generate frequency properties from ingredients that do not have such properties. Such posteriors can, however, be examined on their own behavior, to see if they give sensible results. And also if a prior is being proposed because of the decision-theoretic argument that an optimal procedure is obtained from some choice of prior, then again the posteriors can be examined on their own behavior.

Now consider some distribution $\pi(\theta|y)$, perhaps Bayesian or other, that has been proposed for an unknown parameter value θ in a context where the y value comes from a model $f(y;\theta)$. From the proposed distribution, we can calculate the nominal probability left or right of any value θ of interest, or inversely for any nominal probability level β we can obtain the parameter value $\hat{\theta}_\beta(y)$ having nominal probability β to say the right of it. We consider such β-quantiles as the focal route for assessing a proposed distribution $\pi(\theta|y)$.

In an application there is a true value θ for the parameter and there is a consequent observed y value. And for any chosen β level of interest, the proposed distribution based on the y value leads to the β-quantile point $\hat{\theta}_\beta(y)$. Is the true parameter value in the interval $(\hat{\theta}_\beta(y), \infty)$? Either it is or it is not, from an oracle viewpoint! Then for any given true value θ, we can calculate the probability whether the downstream y value will produce a true statement concerning the presence of θ in the interval $(\hat{\theta}_\beta(y), \infty)$; let this probability be designated $\beta(\theta)$. If the proposed data-based distribution $\pi(\theta|y)$ is sensible in some communicable sense then one would reasonably hope that the $\beta(\theta)$ just calculated would bear some sensible relation to the target value β.

Consider further the proposed distribution $\pi(\theta|y)$. If the calculated $\beta(\theta)$ is identically equal to the value β, then the proposed distribution has the confidence property developed by Fisher[5] and Neyman;[6] and otherwise there would be discrepancies and the proposal would not have the confidence property. Perhaps, however, for some particular prior $\pi_*(\theta)$, the proposal might give an average $\int \beta(\theta) d\pi_*(\theta)$ that is equal to β; then with that prior there is an averaging of the discrepancies that would ~~give~~ the target value β. We give next some examples where such special advantageous priors cannot exist, and thus do not average out discrepant behavior that is intrinsic to a proposed posterior.

lead to

SOME EXAMPLES

We explore some examples, starting with our original Bayes example and progressing through various

OK

complications that arise with the extended Bayes approach.

Example 1. Consider the location model in Section on *The Innovation* using the extreme value error distribution; see Figure 1(a). With the flat prior $\pi(\theta) = c$ recommended by Bayes, we obtain the posterior

$$\pi(\theta|y^0) = g(y^0 - \theta) \qquad (4)$$

where $g(z) = \exp\{-z - e^{-z}\}$; this is graphed in Figure 1(b). Now suppose we are interested in, for example, the 90% quantile or 90% lower bound for θ. The distribution function for $g(z)$ is $G(z) = \exp\{-e^{-z}\}$ and the quantile inverse is $z_\beta = -\log(-\log\beta)$. For the level $\beta = 90\%$, we then obtain $z_{90\%} = 2.25$ which gives the 90% posterior lower bound

$$\hat{\theta}_{90\%} = y^0 - 2.25 \qquad (5)$$

See Figure 2. This is also the 90% confidence lower bound as discussed in Section on *The Innovation*. And correspondingly it has the property that asserting that the true θ in the 90% interval $(y - 2.25, \infty)$ will always have a track record of being true 90% of the time; this track record applies whether we contemplate a fixed θ, or contemplate a distribution $\pi(\theta)$ as the source of the true θ, or examine any arbitrary sequence of values of θ. The example clarifies the original Bayes[1] proposal: that with linearity the Bayes proposal gives what is now called confidence.

Example 2. Consider a response y that is Normal(θ, σ_0^2) but with a mean θ that is known to be greater than or equal to a value θ_0. This may seem rather artificial but it represents the core statistical problem in the analysis of data in the High Energy Physics community; see, for example, Mandelkern,[7] Fraser et al.,[8] and Heinrich;[9] it also has relevance for the current analysis of data from the Large Hadron Collider in Geneva. For expository convenience here, we use $\theta_0 = 0$.

Now for notation let ϕ and Φ designate the density and distribution functions for the standard normal, and let z_β be the upper β-quantile, that is, $\Phi(z_\beta) = \beta$. The observed likelihood from data y^0 is

$$L^0(\theta) = c\phi(y^0 - \theta) \qquad (6)$$

for $\theta \geq \theta_0 = 0$; see Figure 3(a). The model is translation invariant for all possible θ value but does have the restriction $\theta \geq 0$. We can reasonably use the Bayes flat prior on the possible θ. The resulting posterior distribution is

$$\pi(\theta|y^0) = \Phi^{-1}(y^0)\phi(y^0 - \theta) \qquad (7)$$

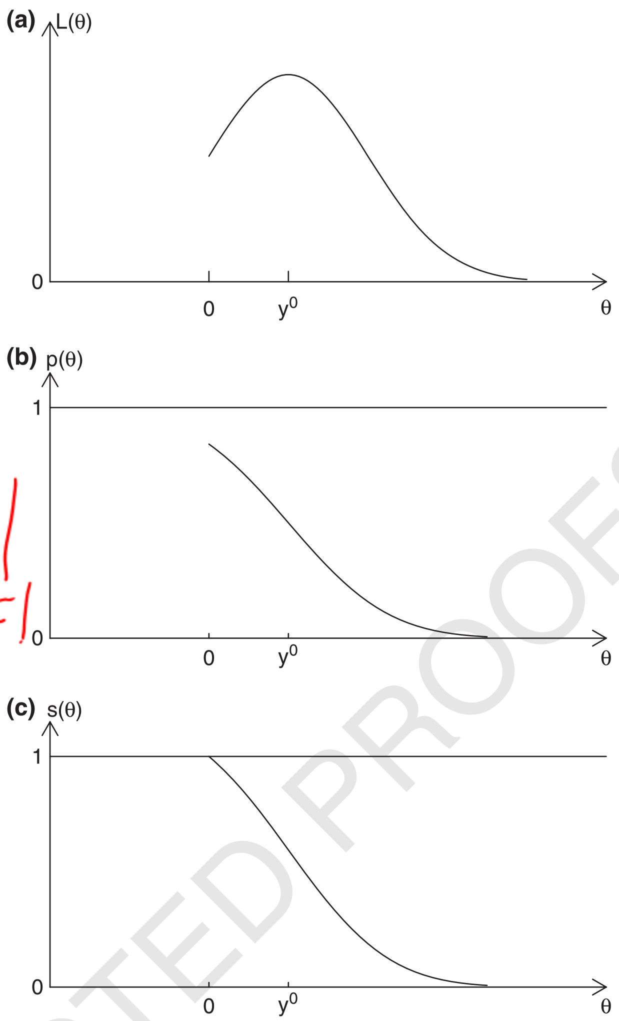

FIGURE 3 | The normal $(\theta, 1)$ with $\theta \geq \theta_0 = 0$: (a) the likelihood function $L(\theta)$; (b) p-value function $p(\theta) = \Phi(y^0 - \theta)$; (c) s-value function $s(\theta) = \Phi(y^0 - \theta)/\Phi(y^0)$.

for $\theta > 0$; the corresponding β-level posterior survivor function is

$$s(\theta) = \Phi^{-1}(y^0)\Phi(y^0 - \theta) \qquad (8)$$

which is plotted in Figure 3(c); and the β-level posterior lower bound is

$$\hat{\theta}_\beta(y^0) = y^0 - z_{\beta\Phi(y^0)} \qquad (9)$$

In order to see some numbers, let us assume $y^0 = 2$. Then if we are interested in the 90%-level posterior lower bound, we obtain

$$\hat{\theta}_{90\%}(2) = 2 - z_{0.880} = 2 - 1.173 = 0.827 \qquad (10)$$

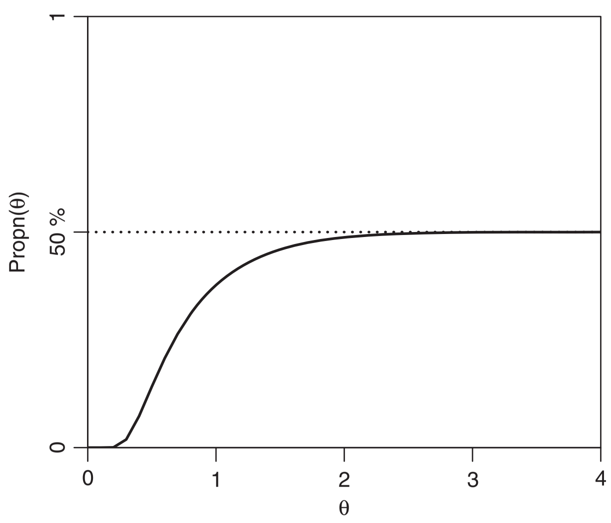

FIGURE 4 | Normal with bounded mean: the actual proportion for the $\beta = 50\%$ quantile is strictly less than 50%.

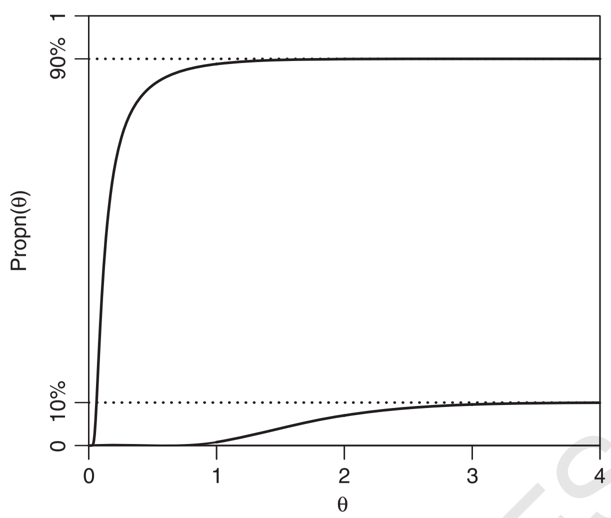

FIGURE 5 | Normal with bounded mean: the actual proportions for the $\beta = 90\%$ and for $\beta = 10\%$ are strictly less than the alleged.

thus with data $y^0 = 2$ we obtain the 90% posterior survivor interval $(0.827, \infty)$. For comparison, we record the 90% confidence lower bound

$$\tilde{\theta}_{90\%}(2) = 2 - 1.282 = 0.718 \qquad (11)$$

with corresponding interval $(0.718, \infty)$. highly

We next examine how the Bayes posterior bound behaves. We are of course concerned with true statements whether the actual θ is in the posterior interval $(\hat{\theta}_\beta(y), \infty)$ calculated at level β. We also comment on certain features associated with the confidence interval $(\tilde{\theta}_\beta(y), \infty)$.

For a particular true value θ, we consider the assertion that the true value is in the Bayes posterior interval $(\hat{\theta}_\beta(y), \infty)$, and calculate the probability that the assertion is true; we obtain

$$\text{Propn}(\theta) = \Pr\{y - z_{\beta\Phi(y)} < \infty; \theta\} = \text{P}\{z < z_{\beta\Phi(\theta+z)}\} \qquad (12)$$

where z is the standard normal. This can be evaluated numerically for particular levels β of interest. In Figure 4, we plot the Propn(θ) for $\beta = 50\%$ and find that the truth of the Bayes statement is uniformly less than the nominal value 50%. Then in Figure 5, we plot the Propn(θ) for $\beta = 90$ and 10%; again the truth of the Bayes statements are uniformly less than the nominal claimed values $\beta = 90$ and 10%. Indeed this holds for all β values; and the results are reversed with for upper bounds.

For the confidence bound $\tilde{\theta}_\beta(y)$, we have in all cases that the true values will be in the interval $(\tilde{\theta}_\beta(y), \infty)$ with probability β. As the calculations are based on where the data point is relative to possible parameter values, we do find that confidence intervals can arise that go beyond the $\theta_0 = 0$ limit. By attempting to avoid this trivial anomaly, the Bayes posterior bound leads to inaccurate results.

Example 3. Consider (y_1, y_2) with a normal distribution located at (θ_1, θ_2) with, for simplicity, the identity covariance matrix. For Bayesian inference, the very natural invariant prior is $\pi(\theta_1, \theta_2) = c$. Then with data $y^0 = (y_1^0, y_2^0)$ the posterior distribution for $\theta = (\theta_1, \theta_2)$ is normal located at (y_1^0, y_2^0) and with again the identity covariance. See Figure 6 for some details concerning the sampling distribution and the posterior distribution.

Now consider a linear parameter $\psi(\theta) = a\theta_1 + b\theta_2$. The marginal posterior distribution is normal with mean $ay_1^0 + by_2^0$ and variance $a^2 + b^2$. The β-level posterior bound is then $\hat{\psi}(y^0) = ay_1^0 + by_2^0 - z_\beta(a^2 + b^2)^{1/2}$. This is also a β-level confidence bound; accordingly if we consider the assertion that the true θ is in the interval $(ay_1^0 + by_2^0 - z_\beta(a^2 + b^2)^{1/2}, \infty)$ we will have that it is true a proportion β of the time, whether the repetitions are with a single true θ, or with an objective prior distribution $\pi(\theta)$ for θ, or with any arbitrary sequence of θ values.● AQ8

Now consider an interest parameter $\rho(\theta) = (\theta_1^2 + \theta_2^2)^{1/2}$ that is nonlinear, ~~for example,~~ quadratic properties ~~as an easy choice~~. In parallel, let $r(y) = (y_1^2 + y_2^2)^{1/2}$ be the natural sample space analog of ρ. Non-central χ^2 distribution theory is immediately

OK

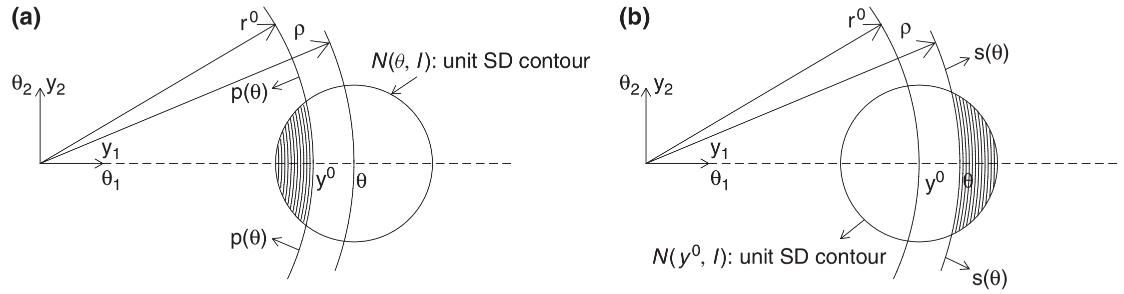

FIGURE 6 | (a) The model is $N(\theta; I)$; region for $p(\theta)$ calculation indicated. (b) The posterior distribution for θ is $N(y^0; I)$; region for $s(\theta)$ calculation indicated.

applicable. The posterior distribution of $\rho(\theta)$ is non-central χ with two degrees of freedom and non-centrality $r(y)$, and we can present this distribution as

$$\rho(\theta) = \{(z_1 + r)^2 + z_2^2\}^{1/2} \tag{13}$$

using standard normal generic variables z_1, z_2. In a similar way, the sampling distribution of $r(y)$ is non-central χ with two degrees of freedom and non-centrality $\rho(\theta)$. Let $G(\chi; \delta)$ be the distribution function of a χ variable with two degrees of freedom and non-centrality δ (see Figure 6).

From the posterior distribution of $\rho(\theta)$, we obtain the β-level posterior lower bound

$$\hat{\theta}_\beta(y) = \chi_{1-\beta}(r) \tag{14}$$

with corresponding β-level interval $\{\chi_{1-\beta}(r), \infty\}$; here $\chi_\alpha(\delta)$ is the solution of $G(\chi; \delta) = \alpha$ for χ. Similarly from the sampling distribution of $r(y)$, we obtain the β-level confidence lower bound

$$\tilde{\theta}_\beta(y) = \delta_\beta(r) \tag{15}$$

with corresponding β-level interval $\{\delta_\beta(r), \infty\}$, where $\delta_\beta(r)$ is the solution of $G(\chi; \delta) = \beta$ for δ.

Now consider the behavior of the intervals in a context where the true value of the full parameter is (θ_1, θ_2). The probability that the true $\rho(\theta)$ is in $(\hat{\theta}_\beta(y), \infty)$ is

$$\begin{aligned} \mathrm{Propn}^{\mathrm{B}}(\theta) &= \Pr\{\hat{\theta}_\beta(y) < \rho(\theta); \theta\} \\ &= \Pr\{\chi_{1-\beta}\{\chi(\rho)\} < \rho; \rho\} \\ &= \Pr\{1 - \beta < G(\rho; \chi(\rho))\} \end{aligned} \tag{16}$$

where $\chi(\rho)$ designates a generic variable having the non-central χ distribution with two degrees of freedom and non-centrality ρ. This is easily available by numerical integration on the real line using R and is plotted in Figure 7 for $\beta = 50\%$ and in Figure 8 for

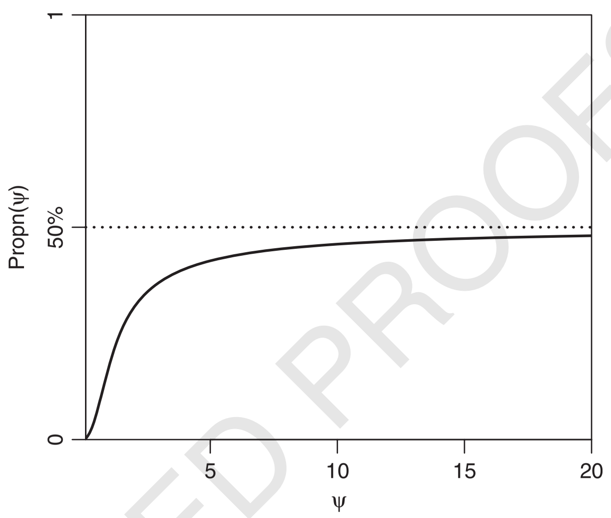

FIGURE 7 | Actual(ρ) proportion with quantile level $\beta = 50\%$.

$\beta = 90$ and 10%. In all cases, the actual Proportion of true statements is uniformly less than the claimed level. In contrast, if the parameter is curved in the other direction, the actual Proportion would be uniformly greater than the claimed.

The behavior of the frequentist interval $\{\tilde{\theta}_\beta(y), \infty\}$ is much easier to evaluate. The probability that the true $\rho(\theta)$ is in the interval $\{\tilde{\theta}_\beta(y), \infty\}$ is

$$\begin{aligned} \mathrm{Propn}^{\mathrm{f}}(\theta) &= \Pr\{\hat{\theta}_\beta(y) < \rho(\theta); \theta\} \\ &= \Pr\{\delta_\beta\{\chi(\rho)\} < \rho; \rho\} \\ &= \Pr\{G(\chi; \rho) < \beta\} \\ &= \beta \end{aligned} \tag{17}$$

And, of course, if θ came from a random source $\pi(\theta)$ then the proportion of cases where the frequentist interval $\{\tilde{\theta}_\beta(y), \infty\}$ includes the true ρ is equal to β; and for any arbitrary sequence of θ values the same is true.

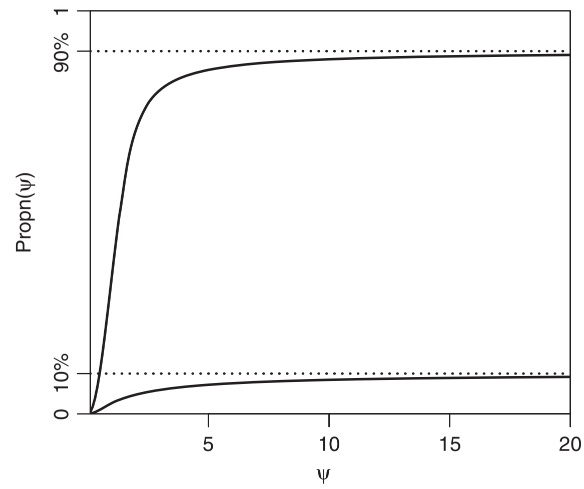

FIGURE 8 | Actual proportion for $\beta = 90$ and 10%.

The complications with a curved parameter represent one aspect of the marginalization paradoxes highlighted by Dawid et al.[10]

Example 4. The preceding examples addressed various departures from the special linearity present in Bayes original proposal. We now examine the simple Normal$(\theta, 1)$ example but slightly modified to allow the variance to depend weakly on the mean θ. Let y be Normal$(\theta, \sigma^2(\theta))$ where $\sigma^2(\theta) = 1 + \gamma\theta^2/n$ and n represents some antecedent sample size that influences the current response distribution. This allows us to use asymptotic theory and focus on the effects of a small presence of model curvature.

2/

$\boxed{\gamma\theta^2/2n}$

The β-quantile of the normal variable y is

$$y_\beta(\theta) = \theta + \sigma(\theta) z_\beta \tag{18}$$

in terms of the standard normal β-quantile z_β. Then working to order $O(n^{-3/2})$ and expanding $(1 + \gamma\theta^2/n)^{1/2}$, we obtain

2/

$$y_\beta(\theta) = \theta + z_\beta \left(1 + \frac{\gamma\theta^2}{4n}\right) \tag{19}$$

The usual confidence inversion then gives the interval $(\tilde{\theta}_\beta(y), \infty)$ where

$$\tilde{\theta}_\beta(y) = y - z_\beta \left\{1 + \frac{\gamma(y - z_\beta)^2}{4n}\right\} \tag{20}$$

A prior examined in standardized coordinates has the asymptotic form

$$\pi(\theta) = c\exp\left\{\frac{a\theta}{n}^{1/2} + \frac{b\theta^2}{2n}\right\} \tag{21}$$

$(1 + \gamma\theta^2/2n)^{1/2}$

$\frac{a\theta}{n^{1/2}}$

A coefficient for θ of order $O(n^{-1/2})$ can easily be seen to over-displace Bayes posterior quantiles. Accordingly, consider priors of the form

hyphen
-|

$$\pi(\theta) = \exp\left\{\frac{a\theta}{n} + \frac{b\theta^2}{2n}\right\} \tag{22}$$

Straightforward calculations then give

$$\hat{\theta}(y) - \tilde{\theta}(y) = (y - z_\beta)\frac{\gamma + c}{2n} + \frac{a}{n} + \frac{cy}{2n} \tag{23}$$

and further calculations give

$$\text{Propn}(\theta) = \beta - \left\{\theta\frac{\gamma}{2n} + \frac{a}{n} + \frac{c}{2n}(2\theta + z_\beta)\right\}\phi(z_\beta) \tag{24}$$

A flat prior in the neighborhood of the center of curvature $\theta = 0$ of the model would have $a = c = 0$. The actual proportion by which the Bayes approach is deficient is then $\theta\gamma\phi(z_\beta)/2n$ which is linear and higher on one side of the center of the model and lower on the other. In fact, it is not possible to choose a prior that gives true statements a claimed level β of the time, unless the prior objectively describes the pattern in which θ values would appear and thereby balance out the positive and negative deficiencies found with the Bayes calculations. For further details, see Fraser.[11]

@

Lindley[12] showed that a confidence distribution in the scalar case could not be a Bayes posterior except in the location model case. This means that accuracy for intervals concerning a parameter cannot in general be obtained by a Bayes calculation from likelihood.

PRIORS FOR BAYESIAN INFERENCE

Preface

The standard statistical context has an observable variable y that is known to be random; the form of the randomness is given by $f(y;\theta)$ for some value of a parameter θ; the true value say $\theta_\star$ is unknown; an observed value y^0 of the variable is obtained from the context which has of course the true value $\theta_\star$; and model information is to be used to make statements concerning what the true value of the parameter is. A statement that the true value is in an interval calculated from the observed y^0 is either true or false and we do not know which.

The confidence approach involves an interval obtained from a y value by a rule or formula; and the confidence level applies over repetitions on the y. If, however, the true $\theta_\star$ is known to have come randomly from some known random source $\pi(\theta)$, then a statement that the true value $\theta_\star$ is in an interval calculated from the observed y^0 is either true or false and we do not know which.

The Bayes approach involves repetitions on the θ value and then subsequently on the y value from that θ; and the posterior level applies over the subset of pairs (θ, y) having $y = y^0$.

With the first approach, the interpretation or meaning is in terms of repeats on y and with the second is in terms of repeats on θ and y then filtered for $y = y^0$.

But if there is not a known source for the true value, then the use of a proposed prior can only be in terms of its success with individual θ values or with various patterns; this was the approach with the examples in the preceding sections. Accordingly we examine various types of prior that can be useful in terms of producing confidence or approximate confidence. We do not address the subjective approach, where $\pi(\theta)$ represents feelings, intuition, elicitation, or guessing.

Location-Based Priors

As described in preceding sections, the original Bayes prior $\pi(\theta) = c$ ~~was proposed~~ is particularly appropriate for a location model $f(y - \theta)$. More general location models can have the form $f(y - X\beta)$ using linear model notation.

An approximation to location properties for more general models was proposed by Jeffreys[13] and made use of Fisher information $I(\theta)$ available for models with moderate regularity:

$$I(\theta) = E\{-\ell_{\theta\theta}(\theta; y); \theta\} \qquad (25)$$

where $\ell_{\theta\theta}(\theta; y) = (\partial/\partial\theta)(\partial/\partial\theta')\log f(y; \theta)$ is the Hessian of the likelihood or log density. The original Jeffreys prior is

$$\pi(\theta)d\theta = |I(\theta)|^{1/2}d\theta \qquad (26)$$

For a scalar parameter, the use of Jefferys prior receives asymptotic support from Welch and Peers.[14] For this, let $\beta(\theta) = \int_{-\infty}^{\theta} I^{1/2}(\theta)d\theta$ be the constant information reparameterization. Then with increasing sample size, we have that $\beta(\hat{\theta}) - \beta(\theta)$ has a fixed distribution to the second order, and thus is pivotal. In effect this is saying that the model is asymptotically location to the second order.

If a model is a location model with scalar or vector parameter, say $f\{a(y) - X\beta(\theta)\}h(y)dy$, then the Jeffreys prior gives

$$f\{a(y) - X\beta(\theta)\}|\frac{\partial\beta}{\partial\theta}|d\theta = f\{a(y) - X\beta(\theta)\}d\beta(\theta) \quad (27)$$

which is effectively a flat prior in the location parameterization.

For the vector parameter case, there are obvious advantages to having a flat prior in the location parameterization, if there is such a parameterization. This was illustrated in Example 3 in the preceding section: probabilities for a linear component parameter had sensible properties; but probabilities for a curved component could be seriously misleading. Thus, some of the resulting probabilities can be acceptable and others can be unacceptable. But this happens with any proposed distribution for a vector parameter: the distribution can only be used sensibly for certain linear types of component parameters: the Dawid et al.[10] marginalization difficulty.

Likelihood-Similar Priors

The location-based priors just discussed were focused on correct statements in quite general circumstances. If we focus on correct statements when there is an acknowledged probability pattern for the θ value that might be present in an application, then we might reasonably use just the corresponding posterior probabilities; but even if the probability pattern is acceptable there can be applications where legal, political, moral, or ethical considerations could limit the use of the probability pattern. Now suppose we do not have an acceptable probability pattern $\pi(\theta)$, and yet want to explore the use of a prior following the approach proposed by Bayes. We could seek some broad class of potential priors that are easy to work with and provide a wide range of function $\pi(\theta)$ to calibrate the parameter. A conjugate class of priors $P = \{\pi(\cdot)\}$ consists of priors $\pi(\theta)$ that have the same functional form as a typical likelihood function and thus are easy to use and to combine with a likelihood function. Consider an example.

Example 1. Suppose $z_1, \ldots, z_n$ is a sample from the Bernoulli(p) distribution. The corresponding likelihood function from data having $\sum z_i = y$ is

$$L(p) = cp^y(1-p)^{n-} \qquad (28)$$

$p^y(1-p)^{n-y}$

A natural conjugate prior family would have

$$\pi(\theta; r, s) = cp^r q^s \qquad (29)$$

with some general range for r and s; these priors are beta distributions for p on $(0, 1)$. The corresponding posterior is $\pi(\theta|y) = cp^{y+r}(1-p)^{n-y+s}$ is then just a beta distribution with modified parameter values. A choice for r and s could be for resulting convenience, or for expediency, or for reasonableness in presenting possibilities concerning the background of the parameter value in the context.

Processing Priors

The location-based priors did not directly address or focus on component parameters that had curvature. The likelihood-similar priors focused on the convenience of combining with likelihood. Another type of prior called a reference prior is based on the processing of prior into posterior, of finding a prior that maximizes a Kullback--Leibler[15] distance from prior to posterior and was initialized by Bernardo[16]; also see Bernardo and Smith.[17]

DISCUSSION

In the usual statistical context with an acceptable statistical model, there is a true value of the parameter say $\theta = \theta_\star$ and there is a consequent observed value $y = y^0$ of the response variable. Interest often then centers on some component parameter say $\psi(\theta)$ of particular interest. And one then wants to make a statement or statements coming from y^0 concerning the true value $\psi(\theta_\star)$, a statement such as the true $\psi(\theta_\star)$ is in an interval $(\hat{\psi}_\beta, \infty)$; for this β presents the reliability say 90%, 95% for the assertion; the calculation of the bound $\hat{\psi}_\beta$ would be based on the data value y^0, on the statistical model, and perhaps on separate proposals as to what the true value $\theta_\star$ might be.

The assessment of any such process would certainly depends on whether the claimed reliability related sensibly to the success rate of the process in making true statements of the form '$\psi(\theta_\star)$ is in $(\hat{\psi}_\beta, \infty)$', in particular, is the success rate equal to β.

If the success rate is to be β regardless of the value of $\theta_\star$, then the process is producing a confidence statement. If the success rate is allowed to fluctuate but to have the average value β, when the true value $\theta_\star$ can be in accord with a prior $\pi_\star(\theta)$, then the process is posterior with respect to that prior pattern.

If a posterior distribution is developed for a vector parameter, then it can have the claimed success rate at most for certain linear component parameters. And if a posterior distribution is developed for a scalar full parameter in a model with curvature, then it needs more than likelihood and it cannot be Bayes. And if a posterior is developed for a parameter based on a particular choice of prior, then discrepancies can average out in the presence of that prior but the success rate can be seriously off-the-target in the presence of other patterns for the possible true parameter value.

Insert here → For a detailed discussion of priors that can lead to statements with the reliability of confidence see Fraser[18].

ACKNOWLEDGEMENT

Very special thanks to Ye Sun and Kexin Ji for extensive help during the preparation of this discussion of Bayes analysis.

REFERENCES

1. Bayes T. An essay towards solving a problem in the doctrine of chances. *Philos Trans R Soc London, Ser A* 1763, 53:370–418; 54:296–325. Reprinted in Biometrika 1958, 45:293–315.
2. Laplace PS. *Théorie Analytique des Probabilités* Paris: Courcier; 1812.
3. Fisher RA. *Statistical Methods and Scientific Inference* London: Oliver and Boyd; 1956.
4. Fisher RA. On the mathematical foundations of theoretical statistics. *Philos Trans R Soc London, Ser A* 1922, 222:309–368.
5. Fisher RA. Inverse probability. *Proc Camb Philol Soc* 1930, 26:528–535.
6. Neyman J. Outline of a theory of statistical estimation based on the classical theory of probability. *Philos Trans R Soc London, Ser A* 1937, 237:333–380.
7. Mandelkern M. Setting confidence intervals for bounded parameters. *Stat Sci* 2002, 17:149–172.
8. Fraser DAS, Reid N, Wong A. Setting confidence intervals for bounded parameters: a different perspective. *Phys Rev D* 2004•, 69:033002. AQ9
9. Heinrich J. The Bayesian approach to setting limits: what to avoid. In: Lyons L, Karagoz UM, eds. *Statistical Problems in Particle Physics, Astrophysics and Cosmology*, Proceedings of PHYSTAT05, London: World Scientific; 2006, 98–102.
10. Dawid AP, Stone M, Zidek JV. Marginalization paradoxes in Bayesian and structural inference. *J R Stat Soc [Ser B]* 1973, 35:189–233.
11. Fraser DAS. *Bayes Posterior is Just a Linear Approximation to Confidence*. Technical Report. Dept. of Statistics, University of Toronto, 2008.

12. Lindley DV. Fiducial distribution and Bayes theorem. *J R Stat Soc [Ser B]* 1958, 20:102–107.
13. Jeffreys H. An invariant form by the prior probabilities in estimation problem. *Proc R Soc [Ser A]* 1946, 186:453–461.
14. Welch BL, Peers HW. On formulae for confidence points based on intervals of weighted likelihood. *J R Stat Soc [Ser A]* 1963, 25:318–329.
15. Kullback S, Leibler RA. On information and sufficiency. *Ann Math Stat* 1951, 22:79–86.
16. Bernardo JM. Reference posterior distributions for Bayesian inference (with discussion). *J R Stat Soc [Ser B]* 1979, 41:113–147.
17. Bernardo JM, Smith AFM. *Bayesian Theory*. Chichester: John Wiley & Sons; 1994.
18. Fraser DAS, Reid N, Marras E, Yi GY. Default priors for Bayesian and frequentist inference. J R Stat Soc (Ser A), 2010, to appear.

FURTHER READING

Stainforth DA, Allen MR, Tredger ER, Smith LA. Confidence, uncertainty and decision-support relevance in climate predictions. *Philos Trans R Soc London, Ser A* 2007, 365: 2145–2162 (See also: Gambling on tomorrow. Modelling the Earth's climate mathematically is hard already. Now a new difficulty is emerging. Economist, August 18, 2007, 69).

QUERIES TO BE ANSWERED BY AUTHOR

IMPORTANT NOTE: Please mark your corrections and answers to these queries directly onto the proof at the relevant place. Do NOT mark your corrections on this query sheet.

Queries from the Copyeditor:

AQ1 Please check whether the affiliation given for the author is ok.
AQ2 Abstract has been taken from the ‘Metadata.xml’ file. Please check if this is ok.
AQ3 Please check whether the sentence ‘And . . . parameter’ conveys the intended meaning.
AQ4 The sentence ’Yet. . .promoted’ seems unclear. Please check and rephrase it if necessary.
AQ5 Please provide the caption for Figure 2.
AQ6 The meaning of the sentence ‘Clearly. . .later’ is not clear. Please check and rephrase it if necessary.
AQ7 A running head short title was not supplied; please check if this one is suitable and, if not, please supply a short title that can be used instead.
AQ8 Please check whether the sentence ‘This is. . .values’ conveys the intended meaning. If not please rephrase the sentence.
AQ9 Please clarify if this number ‘033002’ provided in reference 8 represents page number. If not, please provide the page number for the same.