

Studentization and the determination of p -values

By D.A.S. Fraser and Judith Rousseau

Department of Statistics, University of Toronto, Toronto, Canada M5S 3G3
CEREMADE, University Paris Dauphine and CREST, Paris, France
dfraser@utstat.toronto.edu, rousseau@ceremade.dauphine.fr

ABSTRACT. The original Studentization was the conversion of a sample mean departure into the familiar t -statistic, plus the derivation of the corresponding Student distribution function; the observed value of the distribution function is the observed p -value, as presented in an elemental form. We examine this process in a broadly general context: a null statistical model is available together with observed data; a statistic $t(y)$ has been proposed as a plausible measure of the location of the data relative to what is expected under the null; a modified statistic, say $\tilde{t}(y)$, is developed that is ancillary; the corresponding distribution function is determined, exactly or approximately; and the observed value of the distribution function is the p -value or percentage position of the data with respect to the model.

Such p -values have had extensive coverage in the recent Bayesian literature, with many variations and some preference for two versions labelled p_{ppost} and p_{cpred} . The bootstrap method also directly addresses this Studentization process.

We use recent likelihood theory that gives a third order factorization of a regular statistical model into a marginal density for a full dimensional ancillary and a conditional density for the maximum likelihood variable. The full dimensional ancillary is shown to lead to an explicit determination of the Studentized version $\tilde{t}(y)$ together with a highly accurate approximation to its distribution function; the observed value of the distribution function is the p -value and is available numerically by direct calculation or by Markov chain Monte Carlo or by other simulations.

In this paper, for any given initial or trial test statistic proposed as a location indicator for a data point, we develop: an ancillary based p -value designated p_{anc} ; a special version of the Bayesian p_{cpred} ; and a bootstrap based p -value designated p_{bs} . We then show under moderate regularity that these are equivalent to the third order and have uniqueness as a determination of the statistical location of the data point, as of course derived from the initial location measure. We also show that these p -values have a uniform distribution to third order, as based on calculations in the moderate-deviations region. For implementation the Bayesian and likelihood procedures would perhaps require the same numerical computations, while the bootstrap would require a magnitude more in computation and would perhaps not be accessible. Examples are given to indicate the ease and flexibility of the approach.

Some key words. Ancillary; Bayesian; Bootstrap; Conditioning; Departure measure; Likelihood; p -value; Studentization.

1. INTRODUCTION

Suppose we have a null model $f(y; \theta)$ for a statistical context and wish to judge the acceptability of the model in the presence of data y^0 ; the null model may exist on its own or be a restriction of a larger embedding model. Also suppose we have a statistic $t(y)$ that has been proposed as a plausible measure of departure of data from the model; the statistic may have arisen pragmatically based on physical properties, or may be a simple departure measure for some parameter of an embedding model, such as $t(y) = b(y) - \beta$ where $b(y)$ is say some median type estimate of a parameter say β in the larger model.

It would be quite natural to find that the distribution of $t(y)$ depends on the parameter θ of the null model and thus to want to derive a modified statistic say $\tilde{t}(y)$ that is ancillary, with a θ -free distribution, and yet retains as much as possible of the essential structure of the original statistic. One would also want to obtain the distribution function, say $H(\tilde{t})$, of \tilde{t} so as to calculate the observed p -value, $p^0 = H(\tilde{t}^0)$, as the percentage position of the data with respect to the null model, all in the context of the original proposed statistic $t(y)$. In this formulation the only indications of possible alternatives to the null are found in the choice of the departure statistic $t(y)$. We do not here address this important issue, which may depend heavily on the physical context.

As a simple example, consider a sample allegedly from a normal model with given mean μ_0 plus a proposed departure measure $t(y) = \bar{y}$: we might reasonably hope that the indicated p -value would be $p^0 = H_{n-1}(\tilde{t}^0)$, where H is the Student distribution function and \tilde{t}^0 is the observed value of the t -statistic for assessing y^0 relative to μ_0 ; this p -value is of course the usual Student value recording the percentage position of the data with respect to the normal model located at μ_0 .

The process of developing $\tilde{t}(y)$ from $t(y)$ is here referred to as general Studentization, a generalization of the Student (1908) conversion of $\bar{y} - \mu_0$ into the familiar t -statistic with its Student($n - 1$) distribution function as just described. This problem has had extensive recent discussion in the literature, particularly the Bayesian literature; see for example Bayarri & Berger (2000) and Robins et al (2000).

Frequentist theory gives a simple first order p -value called the plug-in p -value,

$$p_{\text{plug}}^0 = G(t^0; \hat{\theta}^0), \quad (1.1)$$

where $G(t; \theta) = P\{t(y) < t; \theta\}$ is the distribution function for $t(y)$ and the parameter has been replaced by its observed maximum likelihood value. This p -value is well known to be remarkably unreliable in many contexts; see for example, Bayarri & Berger (ibid) and Robins et al (ibid). For the simple normal example the plug-in distribution is that of a sample from a normal distribution with mean μ_0 and standard deviation $\hat{\sigma}^0 = \{\sum(y_i - \mu_0)^2\}^{1/2}$; and the derived distribution of \bar{y} is normal with mean μ_0 and standard deviation $\hat{\sigma}^0/n^{1/2}$; and the resulting plug-in p -value is $p^0 = \Phi(n^{1/2}(\bar{y}^0 - \mu_0)/\hat{\sigma}^0)$; this is centered but underestimates departure from the center. It is of particular interest to note that as a statistic, the preceding p -value is one-one equivalent to the ordinary Student statistic $\tilde{t} = n^{1/2}(\bar{y} - \mu_0)/s_y$. And if the plug-in approach were to be repeated for the modified statistic we would obtain the observed p -value $p^0 = H(\tilde{t}^0)$ where H is the Student(n-1) distribution function and p^0 is the ordinary normal theory p -value.

Bootstrap theory is directly addressing this general Studentization problem; see for example Beran (1988). One samples from the null model distribution using the observed maximum likelihood value for the parameter, and then compounds the process as the double or triple bootstrap. The first order bootstrap is directly the plug-in p -value p_{plug}^0 and typically centers an initial statistic; a double bootstrap can then give appropriate scaling, as indicated for the normal example above; and so on. Indeed, a bootstrap evaluation is a plug-in evaluation.

The recent Bayesian literature, has developed many p -values for this general Studentization problem, but from a different viewpoint. As mentioned in Bayarri & Berger (ibid) “Bayesians have a natural way to eliminate nuisance parameters: integrate them out.” In the present notation with a prior $\pi(\theta)$, the posterior density for y is

$$m(y) = c \int f(y; \theta)\pi(\theta)d\theta$$

with appropriate norming, giving the posterior p -value

$$p_{\text{prior}} = P\{t(y) < t(y^0); m(\cdot)\};$$

see Box (1980). For the simple normal example with the natural flat prior for $\log \sigma$, this direct Bayesian approach produces an improper posterior density $m(y)$ for y . This and other complications, as described in the preceding references, have led the Bayesian approach to seek more refined and incisive methods for using prior densities to obtain p -values, and then to obtain some preference for two versions designated p_{ppost} and p_{cpred} .

In the recent Bayesian approach that yielded p_{cpred} , a posterior density for θ is derived from some aspect of the data designated say $Data_1$,

$$\pi(\theta|Data_1) = cL(\theta; Data_1)\pi(\theta),$$

and then used to eliminate θ from the distribution function say G_2 for $t(y)$ derived from some other aspect of the data say $Data_2$,

$$p^0 = \int G_2(Data_2; \theta)\pi(\theta|Data_1)d\theta.$$

If the full data y^0 is used in both places there is a clear conflict in the probability calculations, often described as double-use of the data. Some obvious difficulties with the double use of data can be avoided by having $Data_1$ in some sense distinct from $Data_2$. Bayarri & Berger (2000) and Robins *et al* (2000) study the case where $Data_1$ is the conditional maximum likelihood estimator given the test statistic $t(y)$; for this, Robins *et al* (2000) show that p_{cpred} is asymptotically uniform to first order, provided $t(y)$ is asymptotically normal.

In many settings however the preceding conditional maximum likelihood estimator is extremely difficult to work with as it can need an explicit expression for the density of $t(y)$, which is often unavailable. Here, following Robert and Rousseau (2003), we take $Data_1$ to be $\hat{\theta}$ and $Data_2$ to be $y|\hat{\theta}$. For this Robert & Rousseau (*ibid*) prove that the resulting p -value p_{cpred} is first-order equivalent to the frequentist p -value $P\{t(y) < t|\hat{\theta}; \theta\}$, for any statistic $t(y)$. Here we accept this p -value as a plausible

contender and examine it with other p -values using recent higher-order likelihood theory.

The recent likelihood theory (for example: Fraser & Reid, 2001; Fraser, 2003) assumes an asymptotic model with p dimensional continuous parameter, moderate regularity, and smoothness of the maximum likelihood statistic $\hat{\theta}(y)$, and to third order gives the existence of an ancillary $a(y)$ of dimension $n - p$ with density $g(a)$ and a corresponding conditional density $h(\hat{\theta}|a, \theta)$ for $\hat{\theta}$ given a ; see Section 2. We thus can think of the model as being $h(\hat{\theta}|a; \theta)g(a)$ on a product space for $(\hat{\theta}, a)$, and accordingly as needed write $t(a, \hat{\theta})$ in place of $t(y)$. Certain technical question arise concerning uniqueness, approximation accuracy, and coordinates for the ancillary; these are discussed in Sections 2 and 3..

The frequentist p -value is obtained from the full ancillary density $g(a)$:

$$p_{\text{anc}}^0 = P_g\{t(a) < t^0\} = G_g(t^0; \hat{\theta}^0) = \int_{t(a; \hat{\theta}^0) < t^0} g(a) da, \quad (1.2)$$

where P_g designates probability using the ancillary density $g(a)$ and $G_g(t; \hat{\theta}^0)$ designates the related distribution function for $t(y)$. We let $\tilde{t}(y)$ designate an asymptotic statistic equivalent to the ancillary p -value function $p_{\text{anc}}(y)$. The distribution of the full ancillary is third-order unique using coordinates specific to the observed value of the maximum likelihood estimate; this raises technical issues that are addressed in Section 3.

We also obtain in Section 2 a Bayesian ancillary distribution $\tilde{g}(a)$ by prior averaging, as for $m(y)$, but restricting the probability being examined to that in a region having $\hat{\theta}$ in a small interval $(\pm\delta/2)$ about the observed maximum likelihood value; we find that this modified Bayesian ancillary is equal to the frequentist ancillary to third order; that is $\tilde{g}(a) = g(a)$. From this we obtain a Bayesian p -value p_B^0 that is equal to the frequentist p -value p_{anc}^0 to third order. We also show that the p_{cpred} proposed by Robert & Rousseau (ibid) is equal to this modified Bayesian p -value and thus to the frequentist ancillary.

For the simple normal example, we will see that the ancillary p -value p_{anc} is just the familiar Student p -value mentioned above and accordingly has good properties.

Also, in the case where the ancillary contours conform to the statistic $t(y)$, that is, $t(a, \hat{\theta}^0) = t(a)$, we have directly from (1.2) that p_{anc} is a true p -value in the sense that it is uniform to the third order. In this paper we prove that this happens generally: that p_{anc} is first order asymptotically uniform for any statistic $t(y)$; that under mild conditions on the asymptotic behaviour of $t(y)$, it is second order asymptotically uniform; and that under stronger asymptotic conditions on $t(y)$, it is third order asymptotically uniform. These asymptotic distributional properties then directly apply to the Bayesian p -value and to the Bootstrap p -value now to be described.

For a bootstrap p -value we let $G_i(t; \theta)$ designate the distribution function for a variable indexed by i as calculated from the model $f(y; \theta)$. Then with p_0 designating some initial function $t(y)$ and with the iteration $p_{i+1} = G_i(p_i; \hat{\theta})$, we have that $p_1 = p_{\text{plug}}$ is the plug in p -value, and that $p_3 = p_{\text{bs}}$ is the proposed bootstrap p -value; it can also be described as a triple plug-in p -value. In Section 4, we show that this bootstrap p -value p_{bs} is third order equivalent to the ancillary p -value p_{anc} , under asymptotic normality of the statistic t and some higher order regularity conditions.

We thus extend Robins et al (2000) in several ways: first by working with a specialized version of p_{cpred} ; second by relaxing the hypotheses on the statistic $t(y)$; and third by obtaining higher order results. The first and second aspects are important as a test statistic can often be complicated with no available asymptotic distribution; see for instance the goodness of fit tests in Robert & Rousseau (2003). Moreover, a p -value provides a universal scale for a test procedure and can be considered from a Bayesian perspective as a calibration of such test procedures (Robert & Rousseau, 2003); thus it is important to be as close as possible to the uniform distribution.

Our results are based on large sample likelihood theory for a continuous model with regularity, and show that general Studentization can be obtained by a frequentist ancillary approach, by a Bayesian ancillary approach, or by a three-level bootstrap approach, and that the results are equivalent to third-order. We also note

that the Bayesian and frequentist p -values are available by direct MCMC simulations while the bootstrap values could require double or triple levels and perhaps not have the same numerical accessibility.

2. THE BAYESIAN AND FREQUENTIST ANCILLARY

Ancillaries, exact and approximate, provide the basis for extending higher order approximation methods to quite general contexts, yielding p -values and marginal likelihoods of high third-order accuracy. These results come from recent theory that gives the existence of a third order ancillary of dimension $n - p$. The labeling of the level surfaces for such an ancillary can be obtained conveniently and intrinsically by using the points of intersection with a surface say

$$S_{\hat{\theta}^0} = \{y : \hat{\theta}(y) = \hat{\theta}^0\};$$

corresponding to a value for the maximum likelihood statistic $\hat{\theta}(y)$, say for convenience the observed value of that statistic. Such a surface has $n - p$ dimensions, is cross sectional to the contours of the ancillary and thus indexes the ancillary contours. Accordingly we take $a = y_c$ where y_c designates a point on the surface $S_{\hat{\theta}^0}$ and let $da = dy_c$ be the Euclidean measure on that surface. Although a third order ancillary does not have uniqueness to third order, we do have that the corresponding density $g(a)$ has such uniqueness (Fraser & Reid, 1995, 2001), up to the labeling of the coordinates.

The simple normal example illustrates some aspects of this, and for notational ease we take the null value μ_0 to be zero. The observed maximum likelihood surface has $\hat{\sigma} = \hat{\sigma}^0$ and is the sphere with Σy_i^2 equal to the observed sum of squares; an obvious ancillary is the unit direction $y/|y|$ but there are many others. If however we record the probability for the ancillary where the ancillary contours intersect a particular maximum likelihood surface we obtain a unique distribution which here is the uniform distribution on the sphere. This unique distribution on a maximum likelihood surface is a general likelihood result and is the basis for the third order p -values and marginal likelihoods.

For the asymptotic model in the new coordinates (Fraser & Reid, 1995) we have

$$f(y; \theta)dy = \exp\{\ell(\theta; y)\}|\ell_{\theta;y}(\hat{\theta}; y)|^{-1}|\hat{j}_{\theta\theta}(\hat{\theta}; y)|dad\hat{\theta}, \quad (2.1)$$

where $\ell(\theta; y)$ is the log-density, $|\ell_{\theta;y}| = |\ell_{\theta;y}(\hat{\theta}; y)\ell'_{\theta;y}(\hat{\theta}; y)|^{1/2}$ is the nominal volume of the p gradient vectors in the $n \times p$ cross Hessian $\ell'_{\theta;y} = (\partial^2/\partial\theta'\partial y)\ell(\theta; y)$, and $\hat{j}_{\theta\theta}$ is the observed information.

Integration over $\hat{\theta}$ using the tangent exponential model extension (Fraser & Reid, 1993) of Barndorff-Nielsen's p^* formula gives the distribution of the ancillary a :

$$g(a)da = \frac{(2\pi)^{p/2}}{e^{c/n}} \exp\{\ell(\hat{\theta}^0; a)\}|\ell_{\theta;y}(\hat{\theta}^0; a)|^{-1}|\hat{j}_{\theta\theta}(\hat{\theta}^0; a)|^{1/2}da.$$

where c is a constant $O(1)$. While the distribution of the ancillary as recorded on any chosen cross-section $S_{\hat{\theta}^0}$ is unique to third order, subject of course to the coordinate labeling, there still can be various ancillaries as noted for the simple normal example. Thus when we write $a(y)$ we are implicitly assuming a particular choice of ancillary and thus a particular linking of points from one maximum likelihood surface to another. This can raise certain technical issues and lead to different parameter inference statements, not of interest here. We do note, however, that with independent scalar coordinates and continuity in the parameter-to-variable relationship the inference issue does not arise, and that for independent vector variables the inference statements can depend on how the parameter is related to the variables, as given typically by pivotal or inverted pivotal quantities.

Now consider a Bayesian ancillary density: we marginalize over the prior and then examine the conditional distribution given the maximum likelihood value:

$$\tilde{g}(a|\hat{\theta}) = \frac{\int_{\Theta} f(a, \hat{\theta}; \theta)\pi(\theta)d\theta}{\int_{\Theta} g(\hat{\theta}; \theta)\pi(\theta)d\theta}.$$

A Laplace integration on the numerator and on the denominator yields

$$\begin{aligned} \tilde{g}(a|\hat{\theta}) &= \frac{\exp\{\ell(\hat{\theta}^0; a)\}|\ell_{\theta;a}(\hat{\theta}^0; a)|^{-1}|\hat{j}_{\theta\theta}(\hat{\theta}^0; a)|^{1/2}e^{H(\hat{\theta}^0)/n}(1 + O_P(n^{-3/2}))}{\int_{S_{\hat{\theta}^0}} \exp\{\ell(\hat{\theta}^0; a)\}|\ell_{\theta;a}(\hat{\theta}^0; a)|^{-1}|\hat{j}_{\theta\theta}(\hat{\theta}^0; a)|^{1/2}e^{H(\hat{\theta}^0)/n}da(1 + O_P(n^{-3/2}))} \\ &= g(a)(1 + O_P(n^{-3/2})); \end{aligned}$$

in other words a Bayesian evaluation of the conditional distribution on the maximum likelihood surface reproduces the frequentist ancillary distribution.

We thus have the significant result that the Bayesian averaging of probability in an interval region about the observed maximum likelihood surface generates a distribution $\tilde{g}(a)$ on that surface and that it is equal to the ancillary distribution on that surface. Or from a somewhat different perspective we can use the prior $\pi(\theta)$ to average the distribution for y and then calculate the distribution of y on the section $\hat{\theta} = \hat{\theta}^0$. In essence and without loss we have interchanged the order in which we do a $\hat{\theta}$ -sectioning and a θ -marginalization. Either way we obtain the distribution $g(a) = \tilde{g}(a)$, which is the marginal distribution of the ancillary statistic a .

Now consider the Bayesian p -value proposed by Robert & Rousseau (2003). The posterior distribution from the marginal for $\hat{\theta}$ at the observed $\hat{\theta}^0$,

$$\pi(\theta|\hat{\theta}^0) = cf(\hat{\theta}^0; \theta)\pi(\theta),$$

is combined with the conditional distribution for $y|\hat{\theta}^0$, producing $\pi(\theta)f(y^0; \theta)$; this is then averaged over θ which as we have noted just gives $c\tilde{g}(a)$. We thus have that the proposed modified Bayesian p_{cpred} ,

$$p_{\text{cpred}}^0 = \int_{\Theta} P \left\{ t(y) < t^0 | \hat{\theta}^0; \theta \right\} \pi(\theta|\hat{\theta}^0) d\theta = \int_{t(a, \hat{\theta}^0) < t^0} \tilde{g}(a|\hat{\theta}^0) da,$$

is equal to the ancillary and direct Bayesian p -values, to third order.

3. THE BAYESIAN-FREQUENTIST p -VALUE IS ASYMPTOTICALLY UNIFORM

3.1. The effective statistic. An observed data value y^0 leads to a maximum likelihood value $\hat{\theta}^0$ and a related maximum likelihood surface $S_{\hat{\theta}^0}$. On this surface there is a unique third-order ancillary density that is typically easy to use ; an integration for $t(y) < t(y^0)$ then gives the observed p -value, say p^0 . There is however no immediate assurance with a particular ancillary that contours of $t(y)$ on $S_{\hat{\theta}^0}$ will correspond to those on other maximum likelihood surfaces. In this section we develop a correspondence and then show that the ancillary p -value is third order uniform, under mild assumptions : the statistic $t(y)$ is assumed to have some regularity in addition to the regularity assumptions made on the model.

Let $a(y)$ be a particular third order ancillary that is indexed by points on some initial maximum likelihood surface S_0 , say $S_{\hat{\theta}^0}$. We define a modified statistic $\bar{t}(y)$

to be equal to $t(y)$ on S_0 and otherwise to be constant on contours of the particular ancillary; thus $\bar{t}(a, \hat{\theta}) = t(a, \hat{\theta}^0) = \bar{t}(a)$. We have then

$$P[\bar{t}(y) < t^0; \theta] = P_g[\bar{t}(a, \hat{\theta}^0) < t^0] = p^0. \quad (3.1)$$

The modification $\bar{t}(y)$ however depends on the particular choice of ancillary and on the coordinates provided by the initial maximum likelihood surface. More specifically, the related cylinder set associated with the \bar{t} and p^0 is defined by $T^0 = \{y = (a, \hat{\theta}); \bar{t}(a) < t^0\}$; on other maximum likelihood surfaces the $\bar{t}(y)$ partition may not agree with the $t(y)$ partition. In order to prove that p^0 is uniform we must link these partitions. For this we develop specialized notation that facilitates this and leads to an effective statistic $\tilde{t}(y)$ corresponding to the use of the ancillary p -value.

First we write the ancillary a to order $n^{-3/2}$ in terms of (\bar{t}, d) and then use the recent likelihood theory to examine $\bar{t}(y)$ conditionally given d .

Let $f(y; \theta, \gamma)$ be an embedding model with an additional scalar parameter γ obtained from the initial model by exponential tilting $e^{\gamma t(y)}$ or by translation in a gradient direction of $t(y)$. This augmented model in effect changes the ancillary from $a(y)$ with dimension $n - p$ to say $d(y)$ with dimension $n - p - 1$, and from regularity we then have that a is a 1-1 function of (\bar{t}, d) . Also the asymptotic theory (for example, Fraser & Reid, 1993, 2001) shows that the dependence of the conditional model on the ancillary can be described or parameterized by a finite number say k of characteristics of the ancillary. Thus for the analysis we can marginalize over the unneeded characteristics and take the effective dimensions for $\hat{\theta}, \bar{t}, d$ to be $p, 1, k - 1$ respectively, with the total dimension now fixed.

Consider the simple normal example with $\mu_0 = 0$: the augmented model by tilting with respect to \bar{y} is just that of a sample from the Normal (μ, σ^2) ; the statistic $d(y)$ corresponds to the location-scale standardized residual and has no effect on the conditional model; $\bar{t}(y) = \bar{y}/s$; and $(\hat{\sigma}, \bar{y}, d)$ is equivalent to the initial y .

We now construct the effective statistic $\tilde{t}(y)$ which links the different \bar{t} values from one maximum likelihood surface to other maximum likelihood surfaces.

An initial observed y^0 gives a maximum likelihood surface $S_0 = S_{\hat{\theta}^0}$ together with observed values $t^0 = \bar{t}^0$ and $d = d^0$. The related cylinder set can be expressed as

$$T^0 = \{(\hat{\theta}, \bar{t}, d); \bar{t} < t^0\};$$

it has probability content $p^0 = p_{\text{anc}}^0$ defined by (1.2), and is θ -free. We now construct a statistic $\tilde{t}(y)$ so that on any maximum likelihood surface it has the same contours as $t(y)$ but typically not the same numerical values on those contours: more specifically for y on $S_{\hat{\theta}}$, we compute the corresponding ancillary p -value $G_g(t(y); \hat{\theta})$, then seek the contour on S_0 with the same p -value, then calculate the corresponding t value, and attribute it to $\tilde{t}(y)$; thus

$$\tilde{t}(y) = G_g^{-1} \left(G_g(t(y); \hat{\theta}); \hat{\theta}^0 \right).$$

3.2. Example. Consider the regression model $y = X\beta + \sigma z$, where $z \sim \mathcal{N}(0, I)$ in \mathbb{R}^n , I is the identity matrix in \mathbb{R}^n and X is the design matrix with full column rank r . Let $t(y) = x'_{r+1}y$ be a suggested test statistic, with x_{r+1} linearly independent of X and thus not in the span $\mathcal{L}(X)$ of the vectors X . The maximum likelihood value is then given by $(\hat{\beta}, \hat{\sigma}) = (b, s/n^{1/2})$ where b is the least square estimate and $s^2 = \sum_i (y_i - \hat{y}_i)^2$ is the sum of squares of residuals. Also let \hat{z} be the residual standardized by the length s . The observed maximum likelihood surface is $S_{\hat{\theta}^0} = \{Xb^0 + s^0\hat{z}; \hat{z} \in \mathcal{S}_0\}$ where \mathcal{S}_0 is the unit sphere in the $n-r$ dimensional space $\mathcal{L}^\perp(X)$ orthogonal to the span of X ; the likelihood surface is then given as $Xb^0 + s^0\mathcal{S}_0$. From normal distribution symmetry this gives that the distribution of the ancillary as recorded on the maximum likelihood surface is uniform with respect to surface volume on the sphere and correspondingly uniform relative to surface volume on \mathcal{S}_0 . Any contour of the test statistic $t(y)$, say $\{y; x'_{r+1}y = t\}$ for fixed t that intersects $S_{\hat{\theta}^0}$ will do so in a sphere of one less dimension and divide the initial sphere into two caps. Let p^0 be the surface volume of the cap corresponding to $\{t(y) < t^0\}$ taken as a proportion of the surface volume of the full sphere. The modified t statistic $\tilde{t}(y)$ can be expressed as $\tilde{t}(y) = \tilde{x}'_{r+1}(Xb^0 + s^0\hat{z})$, where \tilde{x}_{r+1} is the orthogonal projection of x_{r+1} on $\mathcal{L}^\perp(X)$. The set T^0 can then be expressed as

$$T^0 = \{y; \tilde{x}'_{r+1}y/s < \tilde{x}'_{r+1}y^0/s^0\},$$

and $\tilde{t}(y) = \bar{t}(y) = \tilde{x}'_{r+1}y/s$, which is equivalent to the usual Student statistic for testing regression on x_{r+1} after eliminating regression on X .

3.3. Asymptotic uniformity. For ease of notation we work with scalar $\hat{\theta}$ and d but the calculations extend directly to the vector case. We have assumed that $t(y)$ and $(\hat{\theta}, \bar{t}, d)$ are regular in the sense that they have an asymptotic normal distribution with expansions as discussed in Cakmak *et al.* (1994, 1998), and we also have assumed that $G_g(t; \hat{\theta})$ is continuously differentiable in $(t, \hat{\theta})$, with positive density at $\hat{\theta}^0$. This implies that $\tilde{t}(y)$ is also asymptotically normally distributed and is differentiable with higher order expansions. For notational convenience we use coordinates that are standardized to the particular $\theta = \hat{\theta}^0$; accordingly we have under $\theta = \hat{\theta}^0$ that $(\hat{\theta}, \bar{t}, d)$ is first order standard normal and that $g(a) = g(\bar{t}, d)$ is also first order standard normal.

We use the modified notation to examine the probability difference between p^0 and the p -value associated with \tilde{t} .

$$\Delta_a = P[\tilde{t}(y) < t^0; \theta_0] - P[\bar{t}(y) < t^0; \theta_0].$$

The region $\{\tilde{t}(y) < t^0\}$ has boundary $\{\tilde{t}(y) = t^0\}$, here expressed implicitly. We solve the implicit equation and expand \bar{t} around $(0, 0)$ as a function of $(\hat{\theta}, d)$: $\bar{t} = t^0 + b(\hat{\theta}, d)$. The probability difference Δ_a can then be expressed as

$$\Delta_a = \int_d \int_{\hat{\theta}} \left\{ \int_{t^0}^{t^0 + b(\hat{\theta}, d)} f(\hat{\theta}, \bar{t}, d) d\bar{t} \right\} d\hat{\theta} dd, \quad (3.2)$$

where the inner integral gives a positive or negative contribution according to the sign of $b(\hat{\theta}, d)$. At point (i) of the Appendix, we generate an asymptotic expansion for the boundary defining function $b(\hat{\theta}, d)$ and then evaluate the contributions to the integral (3.2). We find that $\Delta_a = 0$ to third order, and thus that $\tilde{t}(y)$ is ancillary and p_{anc} is uniform to third order.

3.4. Asymptotic uniformity under weaker conditions. This completing subsection outlines how uniformity to a lower order may be obtained under weaker assumptions. Robert and Rousseau (2003) prove under relaxed conditions that the special p_{cpred} is asymptotically uniform to first order, whatever the statistic $t(y)$;

thus the same holds for p_{anc} . This robustness property with respect to the test statistic is of wide interest as the test statistic may often be too complicated to yield anything easily concerning its limiting distribution.

We can obtain second order uniformity for p_{cpred} or p_{anc} under somewhat stronger conditions. We do not however require asymptotic normality of the test statistic $t(y)$ as in Robbins et al (2000), but do require the familiar regularity conditions on the model as in Bhattacharya and Ghosh (1978) and Bickel and Ghosh (1990); this provides Laplace and Edgeworth expansions for the posterior and for the maximum likelihood estimator. We have then the following theorem where we let Z_2 be the re-centered and renormalized vector formed from the linearly independent components of the matrix of second derivatives of the log-likelihood:

Theorem 1. *Given standard regularity conditions on the model $f(y; \theta)$ and with a standardized version of $t(y)$ such that $(t, u, Z_2) = \{t(y), \sqrt{n}(\hat{\theta} - \theta), Z_2\}$ converges to a distribution with density h , then p_{cpred} being Uniform $(0, 1)$ to the second order is equivalent to*

$$\int h(t, u, Z_2) \mathbb{I}_{t < G^{-1}(p|u)} [u' z_2 u - \text{tr}\{i^{-1}(\theta_0) z_2\}] dt du dZ_2 = 0 \quad (3.3)$$

for all $p \in [0, 1]$, where $G(t|u)$ is the asymptotic conditional distribution function of $t(y)$ given u , and $i(\theta)$ is the Fisher information matrix.

The proof is outlined at points (ii) and (iii) of the Appendix.

Note that condition (3.3) is satisfied in particular when the limiting distribution is Gaussian; it is also satisfied as soon as t is asymptotically independent of $\hat{\theta}$, even though the limiting distribution might not be Gaussian. Also the methods of the theorem could be adapted to produce the third order results in the preceding section.

4. THE BOOTSTRAP p -VALUE

4.1. Overview. In this Section we show under moderate regularity that the bootstrap applied to a statistic $t(y)$ reduces the distributional dependence on the parameter θ by order $n^{-1/2}$ and thus converts in three stages an initial statistic into a statistic third-order free of the parameter, which is the Bayesian-frequentist p -value.

Again we let $G_i(t; \theta)$ denote the distribution function of an i -th statistic $t_i(y)$ so that $p_{i+1} = G_i(p_i; \hat{\theta})$ is the plug-in modification of p_i ; accordingly we show that p_3 is the Bayesian-frequentist p -value.

4.2. Alternative coordinates. We assume the conditions in Section 3 and for ease of exposition work with the scalar parameter case. Also from Section 3 we have that the modification $\tilde{t}(y)$ of $t(y)$ is third order free of the parameter θ and accordingly we can use the coordinates $(\hat{\theta}, \tilde{t}, d)$ in place of $(\hat{\theta}, \bar{t}, d)$ and then have, with d ancillary, that the bootstrap process can be examined conditionally given d and thus in effect work with the simplified coordinates $(\hat{\theta}, \tilde{t})$.

For a particular step in the bootstrap process with resampling from a current $\hat{\theta}^0$ value we assume that the current variables are standardized so that $(\hat{\theta}, \tilde{t})$ is standard normal and that $\hat{\theta}$ has multiple correlation ρ with $t(y)$ where $|\rho| < 1$. As $t(\hat{\theta}^0, \tilde{t}) = \tilde{t}$, we are able to first order reexpress $t(y)$ as $\tilde{t}(y) + [\rho/(1 - \rho^2)^{1/2}] (\hat{\theta} - \hat{\theta}^0)$.

4.3. First level bootstrap. For a data point y^0 and coordinates relative to the corresponding $\hat{\theta}^0$, the bootstrap distribution of $(\hat{\theta}, \tilde{t})$ is Normal $(0, I)$ to first order. It follows that $t(y)$ is Normal $(0, \gamma^2)$ to first order where $\gamma^2 = 1 + \rho^2/(1 - \rho^2) = (1 - \rho^2)^{-1}$, and thus that the observed p -value based on the bootstrap sample $(\hat{\theta}, \tilde{t})$ is

$$p_1 = P\{t(y) < t^0; \hat{\theta}^0\} = \Phi\{(1 - \rho^2)^{1/2}\tilde{t}\} + O_P(n^{-1/2}). \quad (4.1)$$

As \tilde{t} is Normal $(0, 1)$ it follows that p_1 is first order conservative unless $\rho = 0$. This gives the result in Robins *et al.*(2000), as $\rho = 0$ is equivalent to $t(y)$ having asymptotic mean independent of θ . Thus the first order bootstrap is uniform $(0, 1)$ if and only if $t(y)$ and $\hat{\theta}$ are asymptotically independent. Towards bootstrap results to the second order we now use an asymptotic statistic equivalent to p_1 , which from (4.1) has the form $t_1(y) = \tilde{t} + O_P(n^{-1/2})$.

4.4. Iterated bootstrap. For the the effect of a second bootstrap iteration, we expand t_1 in terms of $\hat{\theta}$ about the current $\hat{\theta}^0$:

$$t_1(y) = \tilde{t}(y) + c_1(\tilde{t})(\hat{\theta} - \hat{\theta}^0)n^{-1/2}. \quad (4.2)$$

To assess the second order effect of the last term we need to use only the first order standard normal distribution for $\hat{\theta}$; thus

$$p_2^0 = P\{t_1(y) < t^0; \hat{\theta}^0\} = P\{\tilde{t}(y) < t^0; \hat{\theta}^0\} = p^0.$$

To obtain results to the next order we work with an asymptotic statistic equivalent to p_2 , which from the preceding has the form $t_2(y) = \tilde{t}(y) + O_P(n^{-1})$. We then expand t_2 as before in terms of $\hat{\theta}$ about the current $\hat{\theta}^0$:

$$t_2(y) = \tilde{t} + c_2(\tilde{t})(\hat{\theta} - \hat{\theta}^0)n^{-1} \quad (4.3)$$

And for the last term we again need only the first order standard normal distribution for $\hat{\theta}$; thus to the third order

$$p_3^0 = P\{t_3(y) < t^0; \hat{\theta}^0\} = P\{\tilde{t}(y) < t^0; \hat{\theta}^0\} = p^0.$$

We do note that these three iterations are needed in general to reach \tilde{t} as the next order term in each iteration has a component of the form $\{\hat{\theta} - \hat{\theta}^0\}^2$ which does not disappear under the standard normal averaging.

5. EXAMPLES AND DISCUSSION

For any full exponential family and for any scalar statistic $t(y)$ not a function of the maximum likelihood statistic, p_{cpred} is equal to the conditional p -value as a consequence of the sufficiency and is thus exactly uniformly distributed. This covers Example 2.2 in Bayarri & Berger (2000) for a sample from the scale exponential using the statistic $t(y) = \min_i y_i$; it also covers the example in Gelman et al (1995) for a sample from the Normal (μ, σ^2) and the same minimum value statistic. It also covers the goodness of fit chi-square test: consider the test against a smooth parametric family $\mathcal{F} = \{f_\theta, \theta \in \Theta\}$, and consider the test statistic $t(y) = \sum_{j=1}^k (N_j - np_j(\hat{\theta}))^2 / (np_j(\hat{\theta}))$ with a fixed number k of bins and N_j observations in the j -th bin.

Simple Taylor expansion around the true value θ implies that

$$\begin{aligned} t(y) &= \sum_{j=1}^k \frac{\left(\sqrt{n}(N_j/n - p_j(\theta)) - \sqrt{n}(\hat{\theta} - \theta)p'_j(\theta)\right)^2}{p_j(\theta)} + O_P(n^{-1/2}) \\ &= \sum_{j=1}^{k-1} W_j^2 \left[\frac{1}{p_j(\theta)} + \frac{1}{p_k(\theta)} \right] + \frac{1}{p_k(\theta)} \sum_{j \neq l} W_j W_l + O_P(n^{-1/2}), \end{aligned}$$

where $W_j = \sqrt{n}(N_j/n - p_j(\theta)) - \sqrt{n}(\hat{\theta} - \theta)p'_j(\theta)$. Asymptotically and conditionally on $u = \sqrt{n}(\hat{\theta} - \theta)$, the vector $W = (W_1, \dots, W_{k-1})$ is distributed as $\mathcal{N}(0, \Omega)$, for some covariance matrix Ω independent of u , as soon as

$$i(\theta) = \int \frac{d_\theta f(y_1; \theta)^2}{f(y_1; \theta)} dy_1 > \sum_{j=1}^k \frac{p'_j(\theta)^2}{p_j(\theta)};$$

otherwise the distribution is degenerate. Therefore t is asymptotically independent on u and p_{cpred} is second order uniform.

ACKNOWLEDGEMENT

The authors express their deep appreciation to Nancy Reid for initiating the present Bayesian-frequentist-bootstrap analysis and for many fruitful discussions of the material. The authors also thank the Editor, an Associate Editor and a referee for very penetrating and helpful comments that led to a substantially improved and more focussed manuscript. The Natural Sciences and Engineering Research Council of Canada has provided support for this research.

APPENDIX

(i) Third order uniformity: proof that $\Delta = 0$. For ease of notation we work with scalar $\hat{\theta}$ and d but the calculations extend directly to the vector case. We have assumed that $t(y)$ and the related $(\hat{\theta}, \bar{t}, d)$ are asymptotic, in the sense that they have an asymptotic normal distribution with expansions as discussed in Cakmak et al (1994, 1998), and that $G_g(t; \hat{\theta})$ is continuously differentiable with respect to $(t, \hat{\theta})$. Then from its definition (4.2) it follows that $\tilde{t}(y)$ is also asymptotic as just described.

Now again for notational convenience we use the standardized coordinates introduced in Section 3 relative to a chosen $\hat{\theta}^0$. Accordingly we have that $(\hat{\theta}, \bar{t}, d)$ is Normal $(0, I)$ to first order, and that (\bar{t}, d) has a standard Normal density to first order.

We expand $b(\hat{\theta}, d)$ in a Taylor series around $(0, 0)$ to order $O(n^{-3/2})$ and make use of $b(0, d) = 0$:

$$\begin{aligned} b(\hat{\theta}, d) &= (a_{10} + a_{11}d/n^{1/2} + a_{12}d^2/2n)\hat{\theta} \\ &+ (a_{20}/n^{1/2} + a_{21}d/n)\hat{\theta}^2/2 + (a_{30}/n)\hat{\theta}^3/6. \end{aligned} \quad (\text{A.1})$$

The definition of the boundary on each surface $S_{\hat{\theta}}$ gives

$$\int_d \left\{ \int_{t^0}^{t^0 + b(\hat{\theta}, d)} g(\bar{t}, d) d\bar{t} \right\} dd = O(n^{-3/2}); \quad (\text{A.2})$$

for each $\hat{\theta}$; thus (A.2) also holds with $b(\theta, d)$ replaced by any one of the terms in (A.1). This eliminates or restricts certain coefficients in (A.1) which can then be written to $O(n^{-3/2})$ as

$$b(\hat{\theta}, d) = \{a_{11}d/n^{1/2} + a_{12}(d^2 - 1)/2n\}\hat{\theta} + (a_{21}d)\hat{\theta}^2/2n;$$

We then have

$$\Delta_a = \int_{\hat{\theta}} \int_d \{b(\hat{\theta}, d) f(\hat{\theta}, t^0, d) + (1/2)b^2(\hat{\theta}, d) f_{\bar{t}}(\hat{\theta}, t^0, d)\} d d d \hat{\theta}$$

where $f_{\bar{t}}(\hat{\theta}, \bar{t}, d)$ designates $(\partial/\partial \bar{t})f(\hat{\theta}, t, d)|_{t^0}$. Integration over the standard normal distribution for d shows zero contribution from the $O(n^{-1})$ terms, and leaves an $O(n^{-1})$ term from $a_{11}d/n^{1/2}$, which in turn leaves no contribution after the integration with respect to the standard normal for $\hat{\theta}$.

(ii) Connection to a conditional p -value. The proof is straightforward and from (2.1) we can write

$$f(y|\hat{\theta}^0; \theta) = \frac{\mathbb{I}_{S_{\hat{\theta}^0}} \exp \ell(\theta; y) |\ell_{\theta; y}|^{-1} |\hat{j}|}{f(\hat{\theta}^0; \theta)},$$

where $f(\hat{\theta}; \theta)$ is the marginal density of $\hat{\theta}$. From (1.2) and then modulating relative to the true θ_0 density we obtain

$$\begin{aligned} p_{\text{cpred}}^0 &= \int_{\Theta} P\{t(y) < t^0 | \hat{\theta}^0; \theta\} \pi(\theta | \hat{\theta}^0) d\theta \\ &= \int_{S_{\hat{\theta}^0}} f(y | \hat{\theta}^0; \theta_0) \mathbb{I}_{t(y) < t^0} f(\hat{\theta}^0; \theta_0) \frac{\int_{\Theta} e^{\ell(\theta; y) - \ell(\theta_0; y)} \pi(\theta) d\theta}{\int_{\Theta} f(\hat{\theta}^0; \theta) \pi(\theta) d\theta} da. \end{aligned} \quad (\text{A.3})$$

Under the conditions in Section 2, we use an Edgeworth expansion (Bhattacharya & Ghosh, 1978) for the density of the maximum likelihood estimator which is uniform in θ together with the usual Laplace expansions as in Section 2 and obtain

$$\int_{\Theta} f(\hat{\theta}^0; \theta) \pi(\theta) d\theta = 1 + O_P(n^{-1}).$$

Then for the numerator integral in (A.3), we use a Laplace expansion of the integral together with an Edgeworth expansion of the density $f(\hat{\theta}; \theta_0)$ and obtain

$$p_{\text{cpred}}^0 = \int_{S_{\hat{\theta}^0}} f(y | \hat{\theta}^0; \theta_0) \mathbb{I}_{t(y) < t^0} A(y) dy_c + R_n n^{-1}, \quad (\text{A.4})$$

with the adjustment factor $A(y)$ given as

$$A(y) = \left\{ 1 + H_1(u_0)/n^{1/2} \right\} \left\{ 1 + \frac{1}{2} \left(u_0' \{ \hat{j} - i(\theta_0) \} u_0 - \text{tr} [i(\theta_0)^{-1} \{ \hat{j} - i(\theta_0) \}] \right) \right\},$$

where $u = n^{1/2}(\theta_0 - \hat{\theta})$, $u_0 = n^{1/2}(\theta_0 - \hat{\theta}^0)$, $H_1(u_0)$ is an odd polynomial function, and $R_n = O_P(1)$. In these calculations we make no assumptions concerning the behaviour of $t(y)$ but we do invoke the usual regularity conditions on the likelihood function. The equation (A.4) shows that p_{cpred} is first order equivalent to the conditional p -value: $P\{t(y) < t^0 | \hat{\theta}^0; \theta_0\} = p_{\hat{\theta}^0; \theta_0}(t^0)$ for any $t(y)$, and is thus uniformly distributed to first order.

(iii) Discrepancy from the conditional p -value. We now examine the discrepancy between p_{cpred} and $p_{\hat{\theta}; \theta_0}(t^0)$:

$$\Delta_c = P[p_{\text{cpred}} < p] - P[p_{\hat{\theta}; \theta_0}(t^0) < p]. \quad (\text{A.5})$$

Consider the expression (A.4) and the form of the adjustment factor $A(y)$ and let $z_2 = n^{1/2}\{j(\theta_0) - i(\theta_0)\}$. Then $\hat{j} - i(\theta_0) = z_2 n^{-1/2} - u_0' \mu_3(\theta_0) n^{-1/2} + O_P(n^{-1})$, where $u_0' \mu_3(\theta_0)$ is the $p \times p$ matrix whose (a, b) component is $\sum_{r=1}^p u_r^0 E_{\theta_0} \{ D_{abr} \log f(X; \theta_0) \}$

and D_{abr} designates the third derivative with respect to the parameter coordinates a, b, r . Let

$$W(t^0, u^0) = \int_{S_{\hat{\theta}^0}} f(y|\hat{\theta}^0, \theta_0) \mathbb{I}_{t(y) < t^0} \{u'_0 z_2 u^0 - \text{tr}[i(\theta_0)^{-1} z_2]\} / 2 dy_c.$$

The calculations at (A.4) then show that

$$p_{\text{cpred}} = p_{\hat{\theta}^0; \theta_0}(t^0) \{1 + H_2(u_0)n^{-1/2}\} + W(t^0, u_0)n^{-1/2} + O_P(n^{-1}),$$

where H_2 is an odd polynomial function.

We now compare $P(p_{\text{cpred}} < p; \theta_0)$ with $p = P(p_{\hat{\theta}^0; \theta_0}(t^0) < p; \theta_0)$. For this we assume that a standardized version of $t(y)$, say $t_s(y)$, has as n goes to infinity a limiting conditional density given $\hat{\theta}^0$ which is positive under the θ_0 distribution. We denote $t_s^0 = t_s(y^0)$ and let $G(\cdot | \hat{\theta}, \theta_0)$ be the distribution function of $t_s(y)$ given $\hat{\theta}$ and θ_0 . We also let E_{θ_0} designate the expectation taken with respect to $f(y; \theta_0)$. We assume that the transformation from $t(y)$ to $t_s(y)$ is monotone increasing, in other words that $\{t(y) < t^0\} \cap S_{\hat{\theta}^0} = \{t_s(y) < t_s^0\} \cap S_{\hat{\theta}^0}$. In the following expression the probabilities are calculated under $f(y; \theta_0)$ and we use the simpler notation $G^{-1}(p|u)$ instead of $G^{-1}(p|u, \theta_0)$ and work up to order $O(n^{-1})$

$$\begin{aligned} \Delta_c &= P(p_{\text{cpred}} < p; \theta_0) - P(p_{\hat{\theta}^0; \theta_0} < p; \theta_0) \\ &= P(G^{-1}(p|u) < t_s < G^{-1}[p\{1 - H_2(u)n^{-1/2}\}] - W_n\{G^{-1}(p|u), u\}n^{-1/2} | u]) \\ &\quad - P(G^{-1}(p|u) \geq t_s^0 > G^{-1}[p\{1 - H_2(u)n^{-1/2}\}] - W_n\{G^{-1}(p|u), u\}n^{-1/2} | u]) \\ &= -E_{\theta_0} [H_2(u)n^{-1/2} - W_n\{G^{-1}(p|u), u\}n^{-1/2}]. \end{aligned}$$

As H_2 is an odd polynomial function in u we find that

$$\Delta_c = (1/2)n^{-1/2} \int f_n(t, u, z_2) \mathbb{I}_{t < G^{-1}(p|u)} [u' z_2 u - \text{tr}\{i^{-1}(\theta_0) z_2\}] dt du dz_2,$$

where $f_n(t, z_2, u)$ is the joint density of $\{t_s(y), z_2, n^{1/2}(\theta_0 - \hat{\theta})\}$. If $f_n(t, z_2, u)$ converges almost surely to a density function $f(t, u, z_2)$, then $\Delta_c = 0$ if and only if

$$\int f(t, u, z_2) \mathbb{I}_{t < G^{-1}(p|u)} [u' z_2 u - \text{tr}\{i^{-1}(\theta_0) z_2\}] dt du dz_2 = 0. \quad (\text{A.6})$$

This completes the proof of Theorem 1.

REFERENCES

Bayarri, M.J. and Berger, J.O. (2000). p -values for composite null models. *J. Amer. Statist. Assoc.* **95**, 1127-1142.

Beran, R.J. (1988). Pre-Pivoting test statistics: a bootstrap view of asymptotic refinements. *J. Amer. Statist. Soc.* **83**, 687-697.

Bhattacharya, and Ghosh, (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6**, 434-451

Bickel, P.J., and Ghosh, J.K. (1990). A Decomposition for the Likelihood Ratio Statistic and the Bartlett Correction - A Bayesian Argument. *Ann. Statist.* **18**, 1070-1090.

Box, G.E.P. (1980). Sampling and Bayes inference in scientific modelling. *J. Roy. Statist. Soc.* **143**, 383-430.

Cakmak, S., Fraser, D.A.S., and Reid, N. (1994). Multivariate asymptotic model: exponential and location approximations. *Utilitas Mathematica* **46**, 21-31.

Cakmak, S., Fraser, D.A.S., McDunnough, P., Reid, N., and Yuan, X. (1998). Likelihood centered asymptotic model: exponential and location model versions. *J. Statist. Planning and Inference* **66**, 211-222.

Fraser, D.A.S. (2003). Likelihood for component parameters. *Biometrika* **90**, 327-339.

Fraser, D.A.S, and Reid, N. (1993). Third order asymptotic models: likelihood functions leading to accurate approximations for distribution functions. *Statist. Sin.* **3**, 67-82.

Fraser, D.A.S., and Reid, N. (1995). Ancillaries and third order significance. *Utilitas Mathematica* **47**, 33-53.

Fraser, D.A.S., and Reid, N. (2001). Ancillary information for statistical inference. In S.E. Ahmed and N. Reid (Eds), *Empirical Bayes and Likelihood Inference*, 185-207. Springer-Verlag.

Gelman, A., Carlin, J.B., Stern, H., and Rubin, D.B. (1995). *Bayesian Data Analysis*, London:Chapman and Hall.

Robert, C., and Rousseau, J. (2003). Bayesian Goodness of Fit - A Distance Approach to Bayesian Goodness of Fit. Preprint

Robins, J.M., van der Vaart, A., and Ventura, V. (2000). The asymptotic distribution of p-values in composite null models. *J. Amer. Statist. Assoc.* **95**, 1143-1156.

Student (1908). The probable error of a mean. *Biometrika* **6**, 1-25.

2006.04.24