

## The Bias in Bayes and How to Measure it

D.A.S. Fraser  
Department of Statistics  
University of Toronto, Toronto  
Ontario, Canada M5S 3G3  
dfraser@utstat.toronto.edu

### Summary

A Bayes prior with a likelihood can give approximate confidence and provide a remarkably flexible approach to statistical inference; but is also known to provide inaccurate perhaps incorrect results. We develop a measure of Bayes bias, first examining a simple Normal model and then progressing to quite general models with scalar and vector parameters. The Bias measure can be interpreted as the lateral displacement of the location standardized likelihood function and thus provides ready access to the effect of a prior on  $p$ -values, confidence bounds, and Bayes posterior bounds. The needed computation is comparable to that for the likelihood function and thus provides an initial option for checking merits of Bayesian computation for high dimensions.

**Some Keywords:** Asymptotics; Bayes bias; High dimensional; Likelihood analysis; Log-model expansion; Regular model; Taylor expansion;

### 1. Introduction

The observed likelihood function from a statistical investigation provides substantial information concerning the value of an unknown parameter in a context being investigated. To obtain the observed likelihood function we need a statistical model in the familiar density form  $f(y; \theta)$ : the likelihood function from data  $y^0$  is  $L(\theta; y^0) = cf(y^0; \theta)$  which is just the natural  $y^0$ -section or observed data-section of the model, and is often used in logarithmic form as  $\ell(\theta; y^0) = \log L(\theta; y^0)$ . A powerful tool then for eliciting information from this observed likelihood function is provided by the Bayesian methodology; this methodology introduces a weight function  $\pi(\theta)$  called a prior that provides a scaling for parameter values and treats  $\pi(\theta)$  as a probability input to the statistical model; with this acquired status for the prior the probability lemma for conditional probability produces a posterior density  $\pi(\theta | y^0) = c\pi(\theta)L(\theta; y^0)$  viewed as a probability description concerning the otherwise unknown parameter value. For some further discussion see Fraser (2011).

The probability lemma has two inputs and for validity each must describe a probability process. For the Bayes application the validity condition holds for the model input,

assuming of course that the model is taken as a valid approximation for data production. But for the prior the validity condition does not hold, when as in most applications it is introduced as a facilitator of the inference process. For example: in the original Bayes (1763) there was additional randomization hypothesized as a conceptional description for the origins of the  $\theta$  value. And in the default Bayes case sometimes called the objective Bayes case the prior is artifact, a mathematical object introduced for inference expedience. And even in cases where a known physical source for the  $\theta$  value is available there is no statistical imperative that says the prior must be combined with the model, and indeed there may be good grounds for keeping the two informations separate thus letting any end-user have the option of combining them if deemed appropriate. Of course, if there is no easy way to extract parameter information from the model with data, then the default Bayes approach has the merit of producing approximate confidence, as demonstrated in Fraser (2011).

In § 2 we consider a simple scalar Normal model and examine the bias coming from the use of a Bayes default prior; we then extend this to the first-order asymptotic model and show that again the Bayes bias is easily calculated and easily interpreted. Then in § 3 we recall the significant Welch-Peers result that the Jeffreys prior with a scalar parameter exponential model converts that model to a location model to second order and thus gives second order equivalence of confidence and Bayes posteriors. Then in § 4 we consider the scalar exponential model and give a general expression for Bayes bias recorded as (2); and in § 5 we extend this to the vector parameter case through the use of a directed location reparameterization (3) obtaining the extended definition of bias as (4). § 6 records some brief discussion.

## 2. The pure case and the first order case

There is merit with most procedures in statistics in knowing how they work in the very simplest of situations. Thus, for measuring the bias in Bayes, we first consider a model that is a scale-standardized Normal  $(\theta; 1)$  and a prior that is log-linear  $\pi(\theta) = c \exp(a\theta)$ . Very simple calculations show that the posterior is  $\pi(\theta | y^0) = c \exp\{-(\theta - y^0 - a)^2 / 2\}$ , which is the Normal  $(y^0 + a; 1)$ . Thus if  $a$  is the slope of the log-prior at the maximum likelihood value  $\hat{\theta}^0$  then the posterior has bias  $B = a$ , a shift  $B$  of the observed likelihood in standardized parameter units. In other words, the posterior is not what is indicated by the likelihood function but is that likelihood, in scale standardized units, shifted laterally by the amount  $B = a$ . We call  $B$  here the bias in Bayes and the bias  $B$  is just the slope  $a$  of the log-prior at the maximum likelihood value  $\hat{\theta}^0$  as appropriately scale standardized; it measures how the prior biases the data information by displacing likelihood. We will see that this is a general phenomenon arising in the use of priors.

Most statistical packages record first order  $p$ -values and first order confidence intervals, typically based on the maximum likelihood value or the score value or the signed likelihood ratio value. These can provide valuable information but are often in mutual conflict except in very simple cases like the preceding Normal example. Nonetheless it seems appropriate to consider Bayes bias initially from a first-order perspective.

For this we need Central Limit type regularity and also need the separation of model information into core information, then first order departure that drops off as  $n^{-1/2}$ , then second order departure that drops off as  $n^{-1}$ , and so on. While this might seem like an idle exercise it has remarkable consequences in allowing or enabling higher order and typically very accurate calculations. In this section we consider just the analysis that separates out the core leaving departures that are first order. Call this first order analysis and it corresponds to just the use of familiar Central Limit type results. Then in the next section we consider the improvements in going to the second order for calculations.

For the first-order analysis of a parameter  $\theta$  we can view the score  $S = \ell_{\theta}(\theta)$  as being Normal  $\{0, j_{\theta\theta}(\hat{\theta})\}$ , the maximum likelihood  $\hat{\theta}$  as being Normal  $\{\theta, j_{\theta\theta}^{-1}(\hat{\theta})\}$ , and the signed likelihood ratio  $r = r(\theta; y)$  as being Normal  $(0, 1)$ . For this, the score  $\ell_{\theta}(\theta) = (\partial / \partial \theta) \ell(\theta)$  is the slope or gradient of log-likelihood at the parameter value of interest,  $j_{\theta\theta}(\hat{\theta})$  is a version of information at the maximum likelihood value, and the signed likelihood root in the scalar  $\theta$  case is  $r = \text{sign}(\hat{\theta} - \theta) [2 \{ \ell(\hat{\theta}) - \ell(\theta) \}]^{1/2}$ ; variations arise depending on the version of information used and where it is evaluated.

Now suppose we examine the Bayes analysis of this first order distribution associated with likelihood. The standardization of the variable and parameter will lead to a standardized prior that is log-linear  $\pi(\theta) = c \exp(a'\theta) / n^{1/2}$  with slope of order  $n^{-1/2}$ ; this will have no first order effect on the posterior.

But then just to check how a non-trivial prior would affect the posterior suppose we consider a prior:  $\pi(\theta) = c \exp(a'\theta)$  targeted on the standardized likelihood function. The log-posterior would then have the following quadratic form in the exponent, ignoring constants:

$$-\frac{1}{2}(\theta - \hat{\theta})' j_{\theta\theta}(\hat{\theta})(\theta - \hat{\theta}) + a'\theta = -\frac{1}{2}\{\theta - \hat{\theta} - j_{\theta\theta}^{-1}(\hat{\theta})a\}' j_{\theta\theta}(\hat{\theta})\{\theta - \hat{\theta} - j_{\theta\theta}^{-1}(\hat{\theta})a\},$$

which implies that the asymptotic posterior distribution of  $\theta$  is Normal  $\{\hat{\theta} + j_{\theta\theta}^{-1}(\hat{\theta})a; j_{\theta\theta}^{-1}(\hat{\theta})\}$ . Thus this first order prior would give a first order posterior with the acquired Bayes bias  $B = j_{\theta\theta}^{-1}(\hat{\theta}^0)a$ .

### 3. The Welch-Peers linearization

For an exponential model with scalar canonical variable and scalar canonical parameter, Welch & Peers (1963) investigated confidence intervals and corresponding Bayes posterior intervals and showed that they are equivalent to the second order, provided the Bayes approach uses the Jeffreys (1946) root-information prior. Thus for the exponential model  $f(y; \theta) = \exp\{s(y)\varphi(\theta) + k(\theta)\}h(y)$  with information  $j_{\varphi\varphi}(\varphi) = -\ell_{\varphi\varphi}(\varphi; y) = -k_{\varphi\varphi}(\theta)$  and relative to the canonical parameter  $\varphi = \varphi(\theta)$  we obtain the prior  $j_{\varphi\varphi}^{1/2}(\varphi)d\varphi$ .

A perhaps more fundamental result implicit in Welch & Peers (1963) is that an exponential model with scalar canonical variable and parameter can be rewritten to second order as a location model. To show this we can integrate the default prior and obtain a location reparameterization

$$\beta(\varphi) = \int_{\varphi}^{\varphi} j_{\varphi\varphi}^{1/2}(\varphi) d\varphi \quad (1)$$

that functions as a location parameter thus allowing the model to be rewritten to second order as a location model  $g(\hat{\beta} - \beta) d\hat{\beta}$ . In fact this extended result is available widely for general asymptotic models with scalar variable and parameter (Cakmak et al., 1998), and makes this version of the root-information prior a widely available procedure for obtaining second order default priors; for further details see Fraser et al.(2012), Fraser et al.(2010b).

A location parameterization such as  $\beta(\varphi)$  has the property that the observed information is constant; thus  $j_{\beta\beta}(\hat{\varphi}) \equiv 1$ . A location parameterization however is not unique and any constant multiple of  $\beta$  is also a location parameter; the advantage of our particular choice  $\beta(\varphi)$  is the built-in standardization to have observed information equal to 1, of course, to second order in moderate deviations.

The role of the location model as an exemplar in inference does deserve more recognition. Let  $g(y - \theta)$  be a scalar variable, scalar parameter location model. Tests and confidence intervals can sometimes be one-sided to the left, or one-sided to the right, or two sided, or occasionally more specialized, depending often on the interests of some particular end user. To avoid this seeming arbitrariness there is merit in defining a basic or fundamental  $p$ -value as probability left-of-the-data or just the observed value of the distribution function. As such the  $p$ -value records the statistical position of the data  $y^0$

with respect to a parameter value  $\theta$ : thus  $p(\theta) = \int_{-\infty}^{y^0} g(y; \theta) dy$  gives a general definition of  $p$ -value. If we now switch and take a parameter viewpoint we would correspondingly think of looking right-of-the-parameter and calculate the posterior survivor value  $s(\theta) = \int_{\theta}^{\infty} \pi(\theta) L(\theta; y^0) d\theta$ . Then, in the special location model case with  $g(y; \theta) = g(y - \theta)$  and with a flat prior  $\pi(\theta) = c$ , trivial calculations show that the two are equal, that is  $p(\theta) = s(\theta)$ , or that Bayes equals frequentist or equivalently that the Bayes procedure has frequentist or repetition validity. The location model thus serves as the primal basis for addressing Bayes repetition validity.

#### 4. Measuring Bayes bias to second order: The scalar case

An exponential model provides full third-order generality for statistical inference in the scalar parameter case (Cakmak et al., 1998) (Andrews et al., 2005). Thus to measure Bayes bias to the second order for a scalar parameter, we restrict attention to an exponential model with canonical parameter  $\varphi(\theta)$  and examine the bias from the Bayes use of a prior  $\pi(\varphi)$ . For this we take the gradient  $a$  of the log-prior at the observed

maximum likelihood value  $\hat{\varphi}^0$ , as in Section 2, but then adjust to obtain the gradient  $B$  with respect to the standardized location parameter. For a given initial prior  $\pi(\varphi)d\varphi$  we first reexpress it in terms of the location parameter  $\beta(\varphi) = \int_{\hat{\varphi}^0}^{\varphi} j_{\varphi\varphi}^{1/2}(\varphi)d\varphi$  obtaining  $\pi(\varphi)j_{\varphi\varphi}^{-1/2}(\varphi)d\beta$  and we then take the log-gradient with respect to the location parameter. This gives the log-gradient

$$\begin{aligned} B &= \frac{\partial}{\partial \beta} \log \{ \pi(\varphi) j_{\varphi\varphi}^{-1/2}(\varphi) \} \big|_{\hat{\varphi}^0} \\ &= \left\{ \frac{\partial}{\partial \beta} \log \pi(\varphi) - \frac{\partial}{\partial \beta} \log j_{\varphi\varphi}^{1/2}(\varphi) \right\} \big|_{\hat{\varphi}^0} \\ &= \left\{ \frac{\partial}{\partial \varphi} \log \pi(\varphi) - \frac{\partial}{\partial \varphi} \log j_{\varphi\varphi}^{1/2}(\varphi) \right\} \big|_{\hat{\varphi}^0} j_{\varphi\varphi}^{-1/2}(\hat{\varphi}^0), \end{aligned} \quad (2)$$

which gives the special log-gradient  $B$  of the prior as the excess of the direct log-gradient over the log-gradient of the Jeffreys prior, all then adjusted to allow intermediate differentiation to take place with respect to the convenient canonical parameter  $\varphi$ . The interpretation is as in Section 2.

*Example 1.* Let  $y$  be Normal  $(m(\theta); \sigma_0^2)$  with data  $y^0$  and a proposed prior  $\pi(\theta)$ . The canonical parameter can always be extracted by sample space differentiation of log-likelihood at the observed data, in an appropriate direction for the variable. We have  $\ell(\theta; y) = -\{y - m(\theta)\}^2 / 2\sigma_0^2$  which gives  $\varphi(\theta) = m(\theta) / \sigma_0$ . The observed maximum likelihood value satisfies  $\ell_{\theta} = \{y^0 - m(\theta)\}m'(\theta) / \sigma_0^2 = 0$  giving  $y^0 = m(\hat{\theta}^0)$ . Then if we want a scaled version of the canonical parameter that has observed information equal to the identity we would replace the initial canonical parameter and use  $\varphi(\theta) = m(\theta) / \sigma_0$ . And then to illustrate (1) we can simplify expressions without loss of generality by using  $\sigma_0 = 1$  and  $m(\theta) = \varphi$  and thus have  $\beta = \varphi$ ; in particular we then have  $j_{\varphi\varphi}^0 \equiv 1$  showing that the model has no curvature as defined in Fraser et al. (2010a).

Now consider the log-gradient for the example and suppose that  $\pi(\theta)$  has been rewritten as  $\pi(\varphi)$  in term of the canonical parameter  $\varphi$ . Then the Bayes bias  $B = \{\partial / \partial \beta\} \log \pi(\varphi) \big|_{\hat{\varphi}^0} = \pi'(\hat{\varphi}^0) / \pi(\hat{\varphi}^0)$ , which in this standardized notation directly describes the effective translation of the standardized Normal limiting distribution that comes from the use of the prior for the analysis.

## 5. Measuring Bayes Bias to second order: the vector case

Now consider an exponential model with vector canonical parameter  $\varphi$ , and observed data  $y^0$  with corresponding maximum likelihood value  $\hat{\varphi}^0$ . Second order inference addresses parameter information in moderate deviations about the observed  $\hat{\varphi}^0$  and as such only involves the second derivatives giving the information matrix  $j_{\varphi\varphi}(\hat{\varphi}^0)$  with

third derivatives of magnitude  $\delta / n^{1/2}$ . For some background on approximate second-order exponential models, see Fraser et al. (2012).

We first examine a directed parameter departure  $\varphi = \hat{\varphi}^0 + \rho u$ , where  $u$  is a unit vector defining the direction of a departure and  $\rho$  is the magnitude of the departure; we then consider the directed information say  $j_{\rho\rho}(\varphi) = u' j_{\varphi\varphi}(\hat{\varphi}^0 + \rho u) u$ . We can consider  $j_{\rho\rho}(\varphi)$  as being the one-dimensional directed information for  $\rho$ , viewed as a scalar parameter  $\rho$  for given  $u$ . We then define a directed location reparameterization  $\beta = \beta(\varphi)$  for given direction  $u$ ,

$$\beta(\varphi) = \int_{\hat{\varphi}^0}^{\varphi} j_{\rho\rho}^{1/2}(\rho) d\rho, \quad (3)$$

and observe that  $\beta$  has constant information with the standardized value 1.

We now apply Section 4 conditionally given the direction  $u$  and calculate the log-gradient vector  $B$  of the prior  $\pi(\varphi)$  with respect to the location linear parameterization  $\beta$  allowing the need to reexpress the prior relative to the location parameter as well as the determination of the resulting gradient with respect to the standardized location parameterization:

$$\begin{aligned} B &= \frac{\partial}{\partial \beta} \log \{ \pi(\hat{\varphi}^0 + \rho u) j_{\varphi\varphi}^{-1/2}(\hat{\varphi}^0 + \rho u) \} \big|_{\hat{\varphi}^0} \\ &= \left\{ \frac{\partial}{\partial \beta} \log \pi(\hat{\varphi}^0 + \rho u) - \frac{\partial}{\partial \beta} \log j_{\varphi\varphi}^{1/2}(\hat{\varphi}^0 + \rho u) \right\} \big|_{\hat{\varphi}^0} \\ &= \left\{ \frac{\partial}{\partial \rho} \log \pi(\hat{\varphi}^0 + \rho u) - \frac{\partial}{\partial \rho} \log j_{\rho\rho}^{1/2}(\hat{\varphi}^0 + \rho u) \right\} \big|_{\hat{\varphi}^0} j_{\rho\rho}^{-1/2}(\hat{\varphi}^0), \end{aligned} \quad (4)$$

which gives the special standardized log-gradient  $B$  of the prior as the excess of the log-gradient with respect to the directed location parameterization  $\beta$  over the log-gradient of the directed Jeffreys prior, all then adjusted to allow intermediate differentiation to take place with respect to the directed canonical parameter  $\varphi$ . The interpretation is as in Section 2.

*Example 2.* Consider the familiar Normal  $(X\beta, \sigma^2 I)$  regression model. The log-model can be written as

$$\log f(y; \theta) = a + \varphi_1' X'y + \varphi_2 y'y - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \beta X'X \beta$$

with canonical parameters  $\varphi_1 = \beta / \sigma^2$  and  $\varphi_2 = -1/2\sigma^2$  and constant  $a$ . For ease of notation we record just for the Normal  $(\mu, \sigma^2)$  case which uses  $X = 1'$ . The negative second derivative of likelihood with respect to  $\varphi = (\varphi_1, \varphi_2)'$  is

$$j_{\varphi\varphi} = n \begin{pmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 2\sigma^4 + 4\mu^2\sigma^2 \end{pmatrix} \quad (5)$$

after reexpression in terms of  $\theta = (\mu, \sigma^2)'$ , and is the covariance matrix of the canonical variables. Now consider observed data  $y^0$  with corresponding maximum likelihood value

$(\hat{\mu}^0, \hat{\sigma}^0)$  or  $\hat{\theta}^0$  in the canonical notation. For departure  $d\theta = (d\mu, d\sigma^2)$  from the maximum likelihood value we can use the Jacobian

$$\partial\varphi\partial(\mu, \sigma^2) = \begin{pmatrix} 1/\sigma^2 & -\mu/\sigma^4 \\ 0 & 1/2\sigma^4 \end{pmatrix}$$

appropriately on either side of (5) to calculate

$$j_{d\theta d\theta} = n \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 1/2\sigma^4 \end{pmatrix}.$$

We now use Welch-Peers in (2) to determine the directed location reparameterizations  $\beta_1$  and  $\beta_2$  corresponding to increments in the directions  $d\mu$  and  $d\sigma^2$

$$\beta_1 = \int_{\hat{\mu}^0}^{\mu} (n/\sigma^2)^{1/2} d\mu = n^{1/2}(\mu - \hat{\mu}^0)\hat{\sigma}^0,$$

$$\beta_2 = \int_{(\hat{\sigma}^0)^2}^{\sigma^2} (n/2\sigma^4)^{1/2} d\sigma^2 = (n/2)^{1/2} \{\log \sigma^2 - \log(\hat{\sigma}^0)^2\};$$

we thus see that  $\beta_1$  in the direction  $d\mu$  is linear in  $\mu$  with  $\sigma^2 = \hat{\sigma}^0$  fixed and  $\beta_2$  in the direction  $d\sigma^2$  is linear in  $\log \sigma^2$  with  $\mu$  fixed; and more generally we have that  $\beta$  is radially linear to second order in the neighbourhood of the data  $y^0$ .

Now consider the Bayes bias in a prior  $\pi(\mu, \sigma^2)$ . The prior  $\pi(\theta)d\theta$  reexpressed locally relative to  $\beta$  is  $c\pi(\mu, \sigma^2)\sigma^2 d\beta_1 d\beta_2$ ; and then to have zero Bayes bias we need a zero log-gradient with respect to the local location variable  $\beta$ . In more detail the differential  $d\mu d\sigma^2 = (2^{1/2}/n\hat{\sigma}^0)\sigma^2 d\beta_1 d\beta_2$  leads to the standardized location-based prior  $\pi(\theta)(2^{1/2}/n\hat{\sigma}^0)\sigma^2$ . Then differentiating the log of this location prior gives the Bayes bias

$$B = \left( \frac{(\partial/\partial\mu)\{\log \pi(\theta) + \log \sigma^2\}}{(\partial/\partial\sigma^2)\{\log \pi(\theta) + \log \sigma^2\}} \right) \Big|_{\hat{\theta}^0} = \left( \frac{(\partial/\partial\mu) \log \pi(\theta)}{(\partial/\partial\sigma^2) \log \pi(\theta) + \sigma^{-2}} \right) \Big|_{\hat{\theta}^0}. \quad (6)$$

## 6. The Welch-Peers linearization

We have examined the effect of a prior on a default Bayes analysis by determining the displacement of the standardized likelihood caused by the presence of the prior. This displacement is the Bayes bias consequent to the introduction of the prior. For an exponential model the technique works with likelihood and thus involves only the dimension of the parameter, and we have illustrated it for parameter dimensions 1 and 2.

For more general regular models with continuous parameters an exponential model approximation, called the tangent exponential model, is needed; for some recent discussion see Reid & Fraser (2010). This exponential model leads to third order analysis and requires sample space differentiation of the log-likelihood function  $\ell(\theta; y)$  at the observed data  $y^0$ ; the differentiation is in sensitivity directions representing how parameter change at the observed maximum likelihood value affects the data variable at the observed data point.

In some cases the sample space derivatives may be inaccessible; the log-likelihood can then be replaced by its sample space mean value using  $\theta = \theta_0$ , and then differentiation with respect to  $\theta_0$  at  $\hat{\theta}^0$  gives the canonical parameter  $\varphi(\theta)$  that determines the tangent exponential model but now just to second order. For further details see Reid & Fraser (2010).

These calculations are of order  $n$  and are thus comparable to the computation of the likelihood function itself; they thus have accessibility whenever the likelihood function is available. This assures that the second order effect of using a prior is readily available without the calculations needed for the Bayes posterior itself.

### Acknowledgement

This paper is based on an address to the International Workshop on Perspectives on High Dimensional Data Analysis held at the Fields Institute in Toronto on June 9, 2011. The Natural Sciences and Engineering Research Council of Canada has provided support for the research.

### References

1. Andrews, D. F., Fraser, D. A. S. & Wong, A. (2005). Computation of distribution functions from likelihood information near observed data. *Journal Statist. Plann. Inference* ,134, 180–193.
2. Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc., London*,53, 370–418.
3. Cakmak, S., Fraser, D. A. S., McDunnough, P., REID, N. & Yuan, X. (1998). Likelihood centered asymptotic model: exponential and location model versions. *J. Statist. Planning and Inference* ,66, 211–222.
4. Fraser, A. M., Fraser, D. A. S. & Fraser, M. J. (2010a). Parameter curvature revisited and the Bayesian frequentist divergence. *Statistical Research: Efron*, 44, 335–346.
5. Fraser, D. A. S. (2011). Is Bayes posterior just quick and dirty confidence? with discussion. *Statistical Science*,26, 299–316.
6. Fraser, D. A. S., Hoang, U., Ji, K., Li, X., Li, L., Lin, W. & Su, J. (2012). Vector exponential models and second order analysis. *Pakistan Journal Statistics Operation Research* , to appear.
7. Fraser, D. A. S., Reid, N., Marras, E. & Yi, G. (2010b). Default priors for Bayesian and frequentist inference. *J. Royal Statist. Soc.* 75, 631–654.
8. Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. A* ,186,453–461.
9. Reid, N. & Fraser, D. A. S. (2010). Mean likelihood and higher order approximations. *Biometrika* ,97, 159–170.
10. Welch, B. L. & Peers, H. W. (1963). On formulae for confidence points based on intervals of weighted likelihoods. *Jour. Royal Statist. Soc. B* ,25, 318–329.