ciuupi: An R package for Computing Confidence Intervals that Utilize Uncertain Prior Information

by Rheanna Mainzer and Paul Kabaila

Abstract We have created the R package **ciuupi** to compute confidence intervals that utilize uncertain prior information in linear regression. Unlike post-model-selection confidence intervals, the confidence interval that utilizes uncertain prior information (CIUUPI) implemented in this package has, to an excellent approximation, coverage probability throughout the parameter space that is very close to the desired minimum coverage probability. Furthermore, when the uncertain prior information is correct, the CIUUPI is, on average, shorter than the standard confidence interval constructed using the full linear regression model. In this paper we provide motivating examples of scenarios where the CIUUPI may be used. We then give a detailed description of this interval and the numerical constrained optimization method implemented in R to obtain it. Lastly, using a real data set as an illustrative example, we show how to use the functions in **ciuupi**.

Introduction

Suppose that $y = X\beta + \varepsilon$ is a random n-vector of responses, X is a known $n \times p$ matrix with linearly independent columns, β is an unknown parameter p-vector and $\varepsilon \sim N(\mathbf{0}, \sigma^2 I_n)$, where σ^2 is unknown. Let a be a specified nonzero p-vector and suppose that the parameter of interest is $\theta = a^\top \beta$. We refer to the model $y = X\beta + \varepsilon$ as the "full" model and the confidence interval for θ based on this model as the "standard" confidence interval. It is often the case that applied statisticians want to utilize uncertain prior information in their analysis. This uncertain prior information may arise from previous experience, scientific background or expert opinion. For example, a common assumption in a factorial experiment is that higher order interaction terms are equal to zero. We consider the uncertain prior information that $\tau = c^\top \beta - t = 0$, where c is a specified nonzero p-vector that is linearly independent of a and t is a specified number. Our interest lies in computing (within a reasonable amount of time) a confidence interval for θ, with minimum coverage probability $1 - \alpha$, that utilizes the uncertain prior information that $\tau = 0$.

One could incorporate uncertain prior information in statistical inference using a Bayesian approach. In other words, a $1 - \alpha$ credible interval for θ could be constructed using an informative prior distribution for τ. However, the **ciuupi** package uses a frequentist approach to utilize the uncertain prior information that $\tau = 0$. Utilizing uncertain prior information in frequentist inference has a distinguished history, which includes Hodges and Lehmann (1952), Pratt (1961), Stein (1962), Cohen (1972), Bickel (1984), Kempthorne (1983, 1987, 1988), Casella and Hwang (1983, 1987), Goutis and Casella (1991), Tseng and Brown (1997), and Efron (2006).

The standard confidence interval has the desired coverage probability throughout the parameter space. However, it does not utilize the uncertain prior information that $\tau = 0$. One may attempt to utilize this uncertain prior information by carrying out a preliminary hypothesis test of the null hypothesis $\tau = 0$ against the alternative hypothesis $\tau \neq 0$. This attempt is based on the following two hopes. Firstly, if the prior information is correct then this test will lead to a confidence interval that is narrower than the standard confidence interval. Secondly, if this prior information happens to be incorrect then this test will effectively lead to the standard confidence interval. Unfortunately, this attempt fails miserably because, for certain values of a, c and X, this post-model-selection confidence interval has minimum coverage probability far below $1 - \alpha$ (see e.g. Kabaila and Giri, 2009b), making it unacceptable.

Kabaila and Giri (2009a) proposed a family of confidence intervals, with minimum coverage probability $1 - \alpha$, that utilize the uncertain prior information that $\tau = 0$ as follows. This family of confidence intervals have expected length that is less than the expected length of the standard interval when the prior information is correct and maximum (over the parameter space) expected length that is not too much larger than the expected length of the standard confidence interval. In addition, these confidence intervals have the same expected length as the standard confidence interval when the data strongly contradict the prior information. The admissibility result of Kabaila et al. (2010) implies that a confidence interval with the desired minimum coverage probability and expected length that is less than that of the standard confidence interval when the prior information is correct, must have an expected length that exceeds that of the standard interval for some parameter values.

Unfortunately, computing these confidence intervals is quite time consuming. Furthermore, there

is no existing R package to compute these confidence intervals. Thus, if one wants to compute the confidence interval proposed by Kabaila and Giri (2009a) and originally computed using MATLAB programs, they may have to write their own programs to do so. The time and skill required to write such programs present large barriers to the use of this confidence interval in applied statistics.

? (**?**, Appendix A) described the family of confidence intervals proposed by Kabaila and Giri (2009a) when σ^2 is known. Each confidence interval in this family is specified by a different tradeoff between its performance when the prior information is correct and its performance when this prior information happens to be incorrect. **?** (**?**) then specified an attractive tradeoff that leads to a unique confidence interval. This interval and its coverage probability and expected length properties can now be easily and quickly computed using the R package **ciuupi**.

This confidence interval has the following three practical applications. Firstly, if σ^2 has been accurately estimated from previous data, as in the factorial experiment example described later, then it may be treated as being effectively known. Secondly, for $n - p$ sufficiently large ($n - p \geq 30$, say), if we replace the assumed known value of σ^2 by its usual estimator in the formula for the confidence interval then the resulting interval has, to a very good approximation, the same coverage probability and expected length properties as when σ^2 is known. Thirdly, some more complicated models (including those considered by **?**, **?**) can be approximated by the linear regression model with σ^2 known when certain unknown parameters are replaced by estimates.

The only information needed to assess the coverage probability and expected length of the confidence interval that utilizes uncertain prior information (CIUUPI) are the values of a, c, X and $1 - \alpha$. We stress that this assessment does not use the observed response y. Indeed, if we want to choose between the CIUUPI and some other confidence interval, such as the standard confidence interval, then this choice must be made prior to any examination of the observed response y.

In this paper we provide motivating examples of scenarios where this confidence interval may be used. We then describe, in detail, the CIUUPI computed by the **ciuupi** package and the numerical constrained optimization method implemented to obtain it. We contrast and compare the CIUUPI with a $1 - \alpha$ credible interval for θ constructed using an informative prior distribution for τ. Lastly, instructions on how to use the functions in **ciuupi** are given, using a real data set, from a factorial experiment, as an illustrative example. We hope that, by making **ciuupi** freely available, statisticians who have uncertain prior information of the type that we specify and wish to utilize it will be encouraged to use the CIUUPI instead of the potentially misleading post-model-selection confidence interval.

Motivating examples

The following motivating examples are provided by Kabaila and Giri (2013). These are examples of scenarios where the **ciuupi** package may be used to find a confidence interval for the parameter of interest θ that utilizes the uncertain prior information that $\tau = 0$.

- Pooling of normal means. Suppose that $y_i = \beta_1 + \varepsilon_i$ for $i = 1, \ldots, n_1$ and $y_i = \beta_1 + \beta_2 + \varepsilon_i$ for $i = n_1 + 1, \ldots, n_1 + n_2$, where the ε_i's are independent and identically distributed (i.i.d.) $N(0, \sigma^2)$. The parameter of interest is $\theta = \beta_1$ and we have uncertain prior information that $\tau = \beta_2 = 0$.

- One-way analysis of variance for two treatments. Suppose that $y_{ij} = \beta_i + \varepsilon_{ij}$ for $i = 1, 2$ and $j = 1, \ldots, n_i$, where the ε_i's are i.i.d. $N(0, \sigma^2)$. The parameter of interest is $\theta = \beta_1$ and we have uncertain prior information that $\tau = \beta_1 - \beta_2 = 0$.

- A 2^k factorial experiment with two or more replicates. The parameter of interest θ is a specified contrast. For factorial experiments it is commonly believed that higher order interactions are negligible. Suppose that the uncertain prior information is that the highest order interaction is zero.

- One-way analysis of covariance with two treatments and normal errors. The parameter of interest θ is the difference in expected responses for the two treatments, for a specified value of the covariate. The uncertain prior information is that the hypothesis of 'parallellism' is satisfied.

- Polynomial regression. Suppose that $y_i = \beta_1 + \beta_2 x_i + \cdots + \beta_p x_i^{p-1} + \varepsilon_i$ for $i = 1, \ldots, n$, where the ε_i's are i.i.d. $N(0, \sigma^2)$. The parameter of interest θ is the expected response for a specified value of the explanatory variable x. The uncertain prior information is that $\tau = \beta_p = 0$.

- Linear regression with at least one interaction term. The parameter of interest θ is a given linear combination of the regression parameters. The uncertain prior information is that a specified interaction term is 0.

In addition to the above examples, **?** have used the **ciuupi** package to aid in the computation of a confidence interval that utilizes uncertain prior information in the following more complicated scenario that arises in the analysis of both clustered and longitudinal data. Suppose that $y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 \overline{x}_i + \eta_i + \varepsilon_{it}$ for $i = 1, \ldots, N$ and $j = 1, \ldots, J$, where $\overline{x}_i = J^{-1} \sum_{j=1}^{J} x_{ij}$, the η_i's are i.i.d. $N(0, \sigma_\eta^2)$, and the ε_{ij}'s are i.i.d. $N(0, \sigma_\varepsilon^2)$. The parameter of interest is $\theta = \beta_1$ and we have uncertain prior information that $\tau = \beta_2 = 0$.

The confidence interval that utilizes uncertain prior information computed by ciuupi

Let $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$, the least squares estimator of $\boldsymbol{\beta}$. Then $\widehat{\theta} = \boldsymbol{a}^\top \widehat{\boldsymbol{\beta}}$ and $\widehat{\tau} = \boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} - t$ are the least squares estimators of θ and τ, respectively. Now let $v_\theta = \text{Var}(\widehat{\theta})/\sigma^2 = \boldsymbol{a}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{a}$ and $v_\tau = \text{Var}(\widehat{\tau})/\sigma^2 = \boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{c}$. The known correlation between $\widehat{\theta}$ and $\widehat{\tau}$ is $\rho = \boldsymbol{a}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{c}/(v_\theta v_\tau)^{1/2}$. Let $\gamma = \tau/(\sigma v_\tau^{1/2})$, a scaled version of τ, and $\widehat{\gamma} = \widehat{\tau}/(\sigma v_\tau^{1/2})$, an estimator of γ. Assume that σ^2 is known.

The confidence interval that utilizes uncertain prior information about τ has the form

$$CI(b, s) = \left[\widehat{\theta} - v_\theta^{1/2} \sigma b(\widehat{\gamma}) - v_\theta^{1/2} \sigma s(\widehat{\gamma}), \ \widehat{\theta} - v_\theta^{1/2} \sigma b(\widehat{\gamma}) + v_\theta^{1/2} \sigma s(\widehat{\gamma}) \right], \tag{1}$$

where $b : \mathbb{R} \to \mathbb{R}$ is an odd continuous function and $s : \mathbb{R} \to \mathbb{R}$ is an even continuous function. In addition, $b(x) = 0$ and $s(x) = z_{1-\alpha/2}$ for all $|x| \geq 6$, where the quantile z_a is defined by $P(Z \leq z_a) = a$ for $Z \sim N(0,1)$. The functions b and s are fully specified by the vector $(b(1), b(2), \ldots, b(5), s(0), s(1), \ldots, s(5))$ as follows. By assumption, $b(0) = 0$, $(b(-1), b(-2), \ldots, b(-5)) = (-b(1), -b(2), \ldots, -b(5))$ and $(s(-1), \ldots, s(-5)) = (s(1), \ldots, s(5))$. The values of $b(x)$ and $s(x)$ for any $x \in [-6, 6]$ are found by cubic spline interpolation for the given values of $b(i)$ and $s(i)$ for $i = -6, -5, \ldots, 0, 1, \ldots, 5, 6$. The functions b and s are computed such that $CI(b, s)$ has minimum coverage probability $1 - \alpha$ and the desired expected length properties. This numerical computation method is described in detail in the next section. Note that the functions b and s are computed assuming that σ^2 is known.

As stated in the introduction, for $n - p$ sufficiently large ($n - p \geq 30$, say), if we replace the assumed known value of σ^2 by $\widehat{\sigma}^2 = (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^\top (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})/(n - p)$ in the formula for $CI(b, s)$ then the resulting interval has, to a very good approximation, the same coverage probability and expected length properties as when σ^2 is known. In **ciuupi**, if no value of σ^2 is supplied then the user is given the option of replacing σ^2 by $\widehat{\sigma}^2$, with a warning that $n - p$ needs to be sufficiently large ($n - p \geq 30$, say).

Numerical constrained optimization method used to compute the vector
$(b(1), b(2), \ldots, b(5), s(0), s(1), \ldots, s(5))$

Let

$$k(x) = \Psi\left(b(x) - s(x), \ b(x) + s(x); \ \rho(x - \gamma), \ 1 - \rho^2 \right)$$

and

$$k^\dagger(x) = \Psi\left(-z_{1-\alpha/2}, \ z_{1-\alpha/2}; \ \rho(x - \gamma), \ 1 - \rho^2 \right),$$

where $\Psi(\ell, u; \mu, \sigma^2) = P(\ell \leq Z \leq u)$ for $Z \sim N(\mu, \sigma^2)$. A computationally convenient expression for the coverage probability of $CI(b, s)$ is

$$CP(\gamma; b, s, \rho) = 1 - \alpha + \int_0^6 \left(k(x) - k^\dagger(x) \right) \phi(x - \gamma) + \left(k(-x) - k^\dagger(-x) \right) \phi(x + \gamma) \, dx, \tag{2}$$

where ϕ denotes the $N(0,1)$ pdf. This coverage probability depends on the unknown parameter γ, the functions b and s, the known correlation ρ and the desired minimum coverage probability $1 - \alpha$. Giri (2008) has shown that $CP(\gamma; b, s, \rho)$ is an even function of γ.

Define the scaled expected length of $CI(b, s)$ to be the expected length of $CI(b, s)$ divided by the expected length of the standard $1 - \alpha$ confidence interval, given by

$$\left[\widehat{\theta} - z_{1-\alpha/2} v_\theta^{1/2} \sigma, \ \widehat{\theta} + z_{1-\alpha/2} v_\theta^{1/2} \sigma \right]. \tag{3}$$

This scaled expected length of $CI(b, s)$ is given by

$$SEL(\gamma; s, \rho) = 1 + \frac{1}{z_{1-\alpha/2}} \int_{-6}^{6} (s(x) - z_{1-\alpha/2}) \, \phi(x - \gamma) \, dx.$$

This scaled expected length depends on the unknown parameter γ, the function s, the known correlation ρ and the desired minimum coverage probability $1 - \alpha$. Giri (2008) has shown that $SEL(\gamma; s, \rho)$ is an even function of γ.

We compute the functions b and s such that $CI(b, s)$ has minimum coverage probability $1 - \alpha$ and the desired expected length properties as follows. For given $\lambda \in [0, \infty)$, we minimize the objective function

$$(SEL(\gamma = 0; s, \rho) - 1) + \lambda \int_{-\infty}^{\infty} (SEL(\gamma; s, \rho) - 1) \, d\gamma, \tag{4}$$

with respect to the vector $(b(1), b(2), \ldots, b(5), s(0), s(1), \ldots, s(5))$, subject to the coverage constraint $CP(\gamma) \geq 1 - \alpha$ for all γ. Equivalently, minimize the objective function

$$\xi(SEL(\gamma = 0; s, \rho) - 1) + (1 - \xi) \int_{-\infty}^{\infty} (SEL(\gamma; s, \rho) - 1) \, d\gamma \tag{5}$$

subject to this constraint, where $\xi = 1/(1 + \lambda)$. A computationally convenient formula for the objective function (4) is

$$\frac{2}{z_{1-\alpha/2}} \int_{0}^{6} (s(h) - z_{1-\alpha/2}) (\lambda + \phi(h)) \, dh. \tag{6}$$

Since we are minimizing this objective function, we can leave out the constant at the front of the integral.

When λ is large, this numerical computation recovers the standard confidence interval (3) for θ. As λ decreases towards 0, this computation puts increasing weight on achieving a small value of $SEL(\gamma = 0; s, \rho)$, i.e. an improved confidence interval performance when the uncertain prior information that $\tau = 0$ is correct. However, as λ decreases, $\max_\gamma SEL(\gamma; s, \rho)$ increases, i.e. the performance of the confidence interval when the prior information happens to be incorrect is degraded. Following **?**, we choose λ such that the "gain" when the prior information is correct, as measured by

$$1 - (SEL(\gamma = 0; s, \rho))^2, \tag{7}$$

is equal to the maximum possible "loss" when the prior information happens to be incorrect, as measured by

$$\left(\max_\gamma SEL(\gamma; s, \rho)\right)^2 - 1. \tag{8}$$

We denote this value of λ by λ^*. Our computational implementation of the constraint $CP(\gamma) \geq 1 - \alpha$ for all γ is to require that $CP(\gamma) \geq 1 - \alpha$ for all $\gamma \in \{0, 0.05, 0.1, \ldots, 8\}$. By specifying constraints on the coverage probability $CP(\gamma)$ for such a fine grid of nonnegative values of γ, we ensure that, to an exceedingly good approximation, $CP(\gamma; b, s, \rho) \geq 1 - \alpha$ for all values of γ.

In summary, we compute the vector $(b(1), b(2), \ldots, b(5), s(0), s(1), \ldots, s(5))$ by minimizing (6), where λ is chosen such that (7) = (8), subject to the constraints $CP(\gamma; b, s, \rho) \geq 1 - \alpha$ for all $\gamma \in \{0, 0.05, 0.1, \ldots, 8\}$. Once $(b(1), b(2), \ldots, b(5), s(0), s(1), \ldots, s(5))$ has been computed in this way, we can easily compute the confidence interval that utilizes the uncertain prior information (CIUUPI) for observed response y.

This constrained optimization procedure is carried out using the slsqp function in the **nloptr** package (see Johnson, 2014). Perhaps surprisingly, the large number of constraints on the coverage probability $CP(\gamma; b, s, \rho)$ is handled well by the slsqp function. The integrals in (2) and (6) are computed as follows. For greater accuracy, each integral is split into a sum of six integrals, with lower and upper endpoints consisting of successive knots. Each of these integrals is then computed using Gauss Legendre quadrature with five nodes. Gauss Legendre quadrature was found to be both faster and more accurate than the R function integrate. This quadrature is carried out using the gauss.quad function in the **statmod** package (see Smyth, 2005).

A comparison of the CIUUPI with a Bayesian interval estimator

Kabaila and Dharmarathne (2015) compare Bayesian and frequentist interval estimators for θ in the linear regression context considered in this paper when σ^2 is unknown. They find that the Bayesian and frequentist interval estimators differ substantially. In this section we compare a $1 - \alpha$ credible interval for θ with the CIUUPI, assuming that σ^2 is known.

For ease of comparison of the CIUUPI with a credible interval, we re-express the regression sampling model as follows. Let the $n \times p$ matrix $\widetilde{X}$ be obtained from X using the transformation described in Appendix B of Kabaila and Dharmarathne (2015). The attractive property of $\widetilde{X}$ is that

$$(\widetilde{X}^\top \widetilde{X})^{-1} = \begin{pmatrix} V & 0 \\ 0 & I_{p-2} \end{pmatrix}, \text{ where } V = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

We re-express the regression sampling model as $\widetilde{y} = \widetilde{X} \left[\vartheta, \gamma, \chi^\top\right]^\top + \widetilde{\varepsilon}$, where $\widetilde{y} = y/\sigma$, $\widetilde{\varepsilon} = \varepsilon/\sigma$ and $\vartheta = \theta/(\sigma v_\theta^{1/2})$. Obviously, $\widetilde{\varepsilon} \sim N(0, I_n)$. Let $(\widehat{\vartheta}, \widehat{\gamma}, \widehat{\chi})$ denote the least squares estimator of $(\vartheta, \gamma, \chi)$. Note that $\widehat{\vartheta} = \widehat{\theta}/(\sigma v_\theta^{1/2})$. Clearly, $(\widehat{\vartheta}, \widehat{\gamma})$ has a bivariate normal distribution with mean (ϑ, γ) and covariance matrix V and, independently, $\widehat{\chi} \sim N(\chi, I_{p-2})$. Dividing the endpoints of the CIUUPI by $\sigma v_\theta^{1/2}$, we obtain the following confidence interval for ϑ:

$$\left[\widehat{\vartheta} - b(\widehat{\gamma}) - s(\widehat{\gamma}), \ \widehat{\vartheta} - b(\widehat{\gamma}) + s(\widehat{\gamma})\right], \tag{9}$$

where the functions b and s have been obtained using the constrained optimization described in the previous section.

The uncertain prior information that $\tau = 0$ implies the uncertain prior information that $\gamma = 0$. The properties of a Bayesian $1 - \alpha$ credible interval depend greatly on the prior distribution chosen for $(\vartheta, \gamma, \chi)$. We have chosen a prior distribution that leads to a credible interval with some similarities to the CIUUPI. Assume that the prior probability density function of $(\vartheta, \gamma, \chi)$ is proportional to $\xi^* \delta(\tau) + (1 - \xi^*)$, where $\xi^* = 1/(1 + \lambda^*)$ and δ denotes the Dirac delta function. In other words, we assume an improper prior density for τ that consists of a mixture of an infinite rectangular unit-height 'slab' and a Dirac delta function 'spike', combined with noninformative prior densities for the other parameters. This prior density is a Bayesian analogue of the weight function used in the weighted average over γ, (5). It may be shown that the marginal posterior density of ϑ is

$$w(\widehat{\gamma}) \, \phi\left(\vartheta; \widehat{\vartheta} - \rho\widehat{\gamma}, 1 - \rho^2\right) + (1 - w(\widehat{\gamma})) \, \phi\left(\vartheta; \widehat{\vartheta}, 1\right), \tag{10}$$

where $w(\widehat{\gamma}) = 1/(1 + \lambda^* \sqrt{2\pi} \exp(\widehat{\gamma}^2/2))$ and $\phi(\cdot; \mu, \nu)$ denotes the $N(\mu, \nu)$ pdf. We note that this posterior density is a mixture of two normal probability density functions, such that the weight given to the posterior density centred at $\widehat{\vartheta}$ increases with increasing $\widehat{\gamma}^2$, when $\lambda^* > 0$. It is evident from (10) that the highest posterior density Bayesian $1 - \alpha$ credible interval may consist of the union of two disjoint intervals. For this reason, we consider the shortest $1 - \alpha$ credible interval.

Note that the graph of the function (10) of ϑ consists of the graph of the function

$$w(\widehat{\gamma}) \, \phi\left(\vartheta; -\rho\widehat{\gamma}, 1 - \rho^2\right) + (1 - w(\widehat{\gamma})) \, \phi\left(\vartheta; 0, 1\right),$$

shifted to the right by $\widehat{\vartheta}$. We can therefore express the shortest $1 - \alpha$ credible interval for ϑ in the form $[\widehat{\vartheta} + l(\widehat{\gamma}), \widehat{\vartheta} + u(\widehat{\gamma})]$, for the appropriate functions l and u. We compare this interval with the frequentist $1 - \alpha$ confidence interval (9) as follows. Let $b_B(\widehat{\gamma}) = -(l(\widehat{\gamma}) + u(\widehat{\gamma}))/2$ and $s_B(\widehat{\gamma}) = (u(\widehat{\gamma}) - l(\widehat{\gamma}))/2$. Then $[\widehat{\vartheta} + l(\widehat{\gamma}), \widehat{\vartheta} + u(\widehat{\gamma})]$ is equal to

$$\left[\widehat{\vartheta} - b_B(\widehat{\gamma}) - s_B(\widehat{\gamma}), \ \widehat{\vartheta} - b_B(\widehat{\gamma}) + s_B(\widehat{\gamma})\right], \tag{11}$$

which has a similar form to (9), but with b and s replaced by b_B and s_B respectively. Therefore, we may compare the interval (9) with (11) by comparing the functions b and s with the functions b_B and s_B, respectively. We will also compare the interval (9) with (11) by comparing the frequentist coverage probability function of (11).

Using the ciuupi package

In this section we use a real data set to illustrate how each of the six functions in **ciuupi** works. Table 1 below gives the name of each of the functions and a short description of what it does. In the following subsections we show how the functions in Table 1 are used in R.

Factorial experiment example

Consider the 2×2 factorial experiment described by Kabaila and Giri (Discussion 5.8, 2009a), which has been extracted from a real 2^3 factorial data set provided by Box et al. (1963). The two factors are the time of addition of HNO_3 and the presence of a 'heel'. These factors are labelled A and B, respectively.

Function	Description
bsciuupi	Compute the optimized vector $(b(1), b(2), \ldots, b(5), s(0), s(1), \ldots, s(5))$
bsspline	Evaluate $b(x)$ and $s(x)$ at x
cpci	Evaluate $CP(\gamma; b, s, \rho)$ at γ
selci	Evaluate $SEL(\gamma; b, s, \rho)$ at γ
ciuupi	Compute the CIUUPI, i.e. compute $CI(b, s)$
cistandard	Compute the standard confidence interval

Table 1: Functions in the **ciuupi** package

Define $x_1 = -1$ and $x_1 = 1$ for "Time of addition of HNO_3" equal to 2 hours and 7 hours, respectively. Also define $x_2 = -1$ and $x_2 = 1$ for "heel absent" and "heel present", respectively. Assume a model of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon,$$

where $\varepsilon \sim N(0, \sigma^2)$. This model can be written in matrix form as

$$y = X\beta + \varepsilon$$

where $\beta = (\beta_0, \beta_1, \beta_2, \beta_{12})$,

$$X = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

and $\varepsilon \sim N(0, \sigma^2 I_n)$. According to Box et al. (1963), a very accurate estimate of σ, obtained from previous related experiments, is 0.8.

Suppose that the parameter of interest θ is (expected response when factor A is high and factor B is low) − (expected response when factor A is low and factor B is low). In other words, $\theta = 2(\beta_1 - \beta_{12})$, so that $\theta = a^\top \beta$, where $a = (0, 2, 0, -2)$. Our aim is to find a confidence interval, with minimum coverage 0.95, for θ. We suppose that there is uncertain prior information that the two-factor interaction is zero. In other words, we suppose that there is uncertain prior information that $\beta_{12} = 0$. The uncertain prior information is, then, that $\tau = c^\top \beta - t = 0$, where $c = (0, 0, 0, 1)$ and $t = 0$. Now that we have specified a, c and X, we can compute $\rho = a^\top (X^\top X)^{-1} c / (v_\theta \, v_\tau)^{1/2} = -1/\sqrt{2} = -0.707$.

Evaluating the confidence interval (no examination of the observed response)

First suppose that we have not yet examined the observed response y and that we are interested in knowing how the confidence interval that utilizes uncertain prior information (CIUUPI) performs for given values of $1 - \alpha$, a, c and X. We begin by storing the values of α, a, c and X in R as follows.

```
# Specify alpha, a, c and x.
alpha <- 0.05
a <- c(0, 2, 0, -2)
c <- c(0, 0, 0, 1)
x <- cbind(rep(1, 4), c(-1, 1, -1, 1), c(-1, -1, 1, 1), c(1, -1, -1, 1))
```

Next we use the numerical constrained optimization to compute the values at the knots of the functions b and s that define the CIUUPI. We must specify whether natural cubic spline interpolation (natural = 1) or clamped cubic spline interpolation (natural = 0) is used in the description of these functions. In the case of clamped cubic spline interpolation the first derivatives of b and s are set to zero at -6 and 6. Natural cubic spline interpolation is the default, and is carried out using splinefun in the **stats** package. Clamped cubic spline interpolation is carried out using cubicspline in the **pracma** package. The nonlinear constrained optimization using natural cubic spline interpolation for the description of the functions b and s is much faster and results in a coverage probability that is slightly closer to $1 - \alpha$ throughout the parameter space. For this example we are able to obtain the vector $(b(1), b(2), \ldots, b(5), s(0), s(1), \ldots, s(5))$ in 6.56 minutes when using natural cubic spline interpolation and in 21.27 minutes when using clamped cubic spline interpolation. This computation was carried out on a PC with an Intel i7-7500 CPU (3.4GHz) and 32GB of RAM. The following code is used to obtain the vector $(b(1), b(2), \ldots, b(5), s(0), s(1), \ldots, s(5))$ that specifies the CIUUPI, which is obtained from the numerical constrained optimization that uses natural cubic spline interpolation for the description of the functions b and s.

```
# Compute (b(1), b(2), ..., b(5), s(0), s(1), ..., s(5)) that specifies the CIUUPI
bsvec <- bsciuupi(alpha, a = a, c = c, x = x)
bsvec
```

Alternatively, since we know that $\rho = -0.707$, we could obtain the vector $(b(1), b(2), \ldots, b(5), s(0), s(1), \ldots, s(5))$ that specifies the CIUUPI using the code

```
# Compute (b(1), b(2), ..., b(5), s(0), s(1), ..., s(5)) that specifies the CIUUPI,
# given rho
bsvec2 <- bsciuupi(alpha, rho = -0.707)
```

Now that we have the vector $(b(1), b(2), \ldots, b(5), s(0), s(1), \ldots, s(5))$ that specifies the CIUUPI, we can graph the functions b and s using the following code:

```
# Compute the functions b and s that specify the CIUUPI on a grid of values
splineval <- bsspline(seq(0, 8, by = 0.1), bsvec, alpha)

# The first 5 values of bsvect are b(1),b(2),...,b(5).
# The last 6 values are s(0),s(1),...,s(5).
xseq <- seq(0, 6, by = 1)
bvec <- c(0, bsvec[1:5], 0)
svec <- c(bsvec[6:11], qnorm(1 - alpha/2))

# Plot the functions b and s
plot(seq(0, 8, by = 0.1), splineval[, 2], type = "l", main = "b function",
    ylab = " ", las = 1, lwd = 2, xaxs = "i", col = "blue", xlab = "x")
points(xseq, bvec, pch = 19, col = "blue")
plot(seq(0, 8, by = 0.1), splineval[, 3], type = "l", main = "s function",
    ylab = " ", las = 1, lwd = 2, xaxs = "i", col = "blue", xlab = "x")
points(xseq, svec, pch = 19, col = "blue")
```

Figure 1 shows the graphs of the functions b and s that specify the CIUUPI, when these functions are described using natural cubic spline interpolation, for this example. For comparison, Figure 2 shows the graphs of the functions b and s that specify the CIUUPI, when these functions are described using clamped cubic spline interpolation. These figures are quite similar; there is a small difference in both the b and s functions near $x = 6$.

We can also use the vector $(b(1), b(2), \ldots, b(5), s(0), s(1), \ldots, s(5))$ that specifies the CIUUPI to evaluate and then plot the coverage probability $CP(\gamma; b, s, \rho)$ and scaled expected length $SEL(\gamma; s, \rho)$ as functions of γ. This is done using the following code.

```
# Compute the coverage probability and scaled expected for a grid of values of gamma
gam <- seq(0, 10, by = 0.1)
cp <- cpciuupi(gam, bsvec, alpha, a = a, c = c, x = x)
sel <- selciuupi(gam, bsvec, alpha, a = a, c = c, x = x)

# Plot the coverage probability and squared scaled expected length
plot(gam, cp, type = "l", lwd = 2, ylab = "", las = 1, xaxs = "i",
main = "Coverage Probability", col = "blue",
xlab = expression(paste("|", gamma, "|")), ylim = c(0.9495, 0.9505))
abline(h = 1-alpha, lty = 2)
plot(gam, sel^2, type = "l", lwd = 2, ylab = "", las = 1, xaxs = "i",
main = "Squared SEL", col = "blue",
xlab = expression(paste("|", gamma, "|")), ylim = c(0.83, 1.17))
abline(h = 1, lty = 2)
```

Figure 3 shows the graphs of $CP(\gamma; b, s, \rho)$ and the square of $SEL(\gamma; b, s, \rho)$ for the CIUUPI (where the functions b and s have been specified by natural cubic spline interpolation) produced by this code.

We can see from Figure 3 that, regardless of the value of γ, the coverage probability of the CIUUPI is extremely close to $1 - \alpha$. We can also see that the expected length of the CIUUPI is less than the expected length of the standard confidence interval when γ is small, with the minimum scaled expected length achieved when $\gamma = 0$. For moderate values of $|\gamma|$, the expected length of the standard interval is less than the expected length of the CIUUPI. However, for large $|\gamma|$, the expected length of the CIUUPI is essentially the same as the expected length of the standard interval.

For comparison, Figure 4 shows the graphs of $CP(\gamma; b, s, \rho)$ and the square of $SEL(\gamma; b, s, \rho)$ for the CIUUPI when the functions b and s are described by clamped cubic spline interpolation.
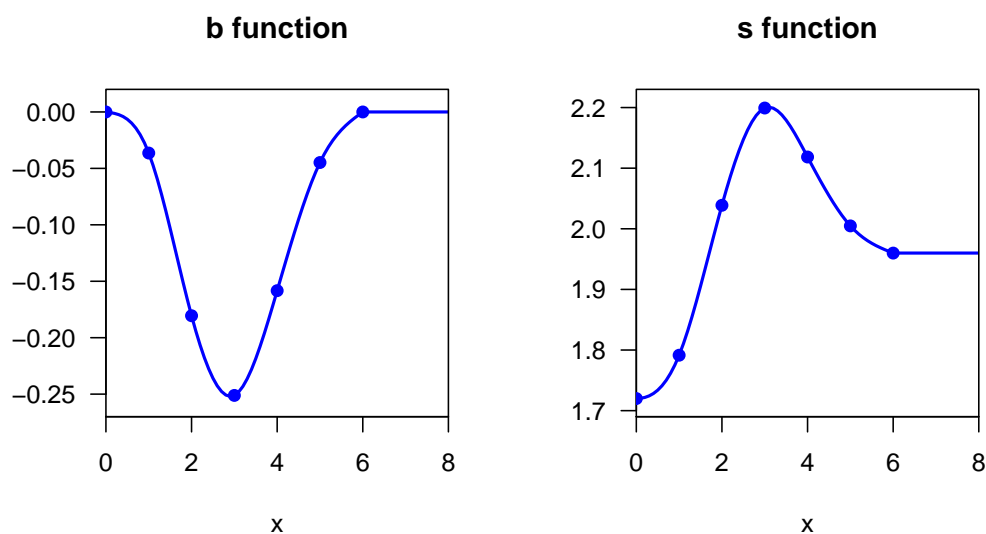
Figure 1: Graphs of the functions b and s for the factorial experiment example for the CIUUPI, with minimum coverage probability 0.95, when they are described using natural cubic spline interpolation, for the factorial experiment example.
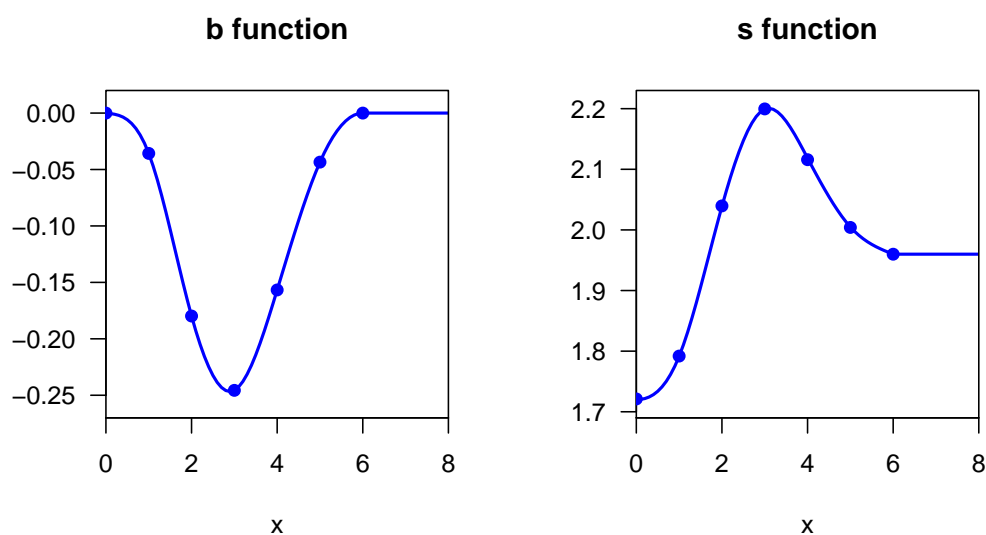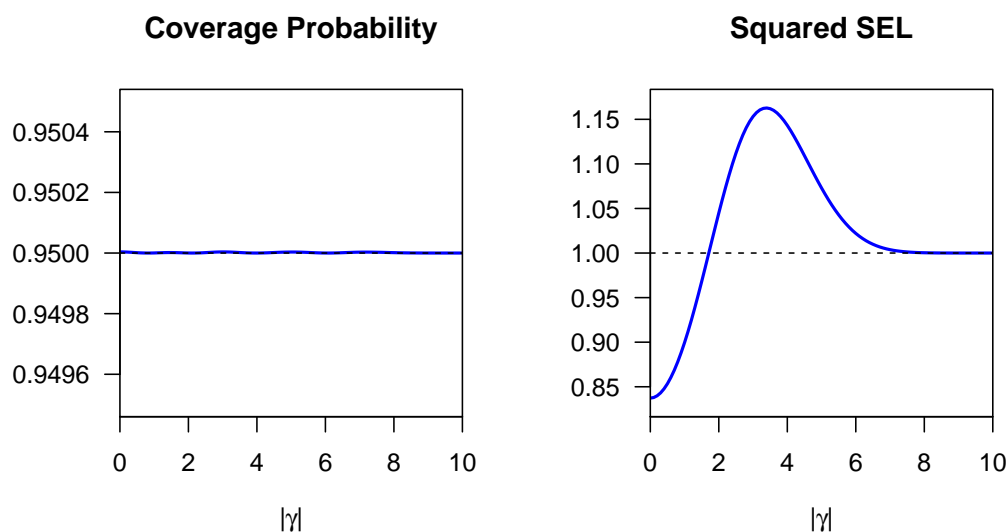

Figure 2: Graphs of the functions b and s for the factorial experiment example for the CIUUPI, with minimum coverage probability 0.95, when they are described using clamped cubic spline interpolation, for the factorial experiment example.

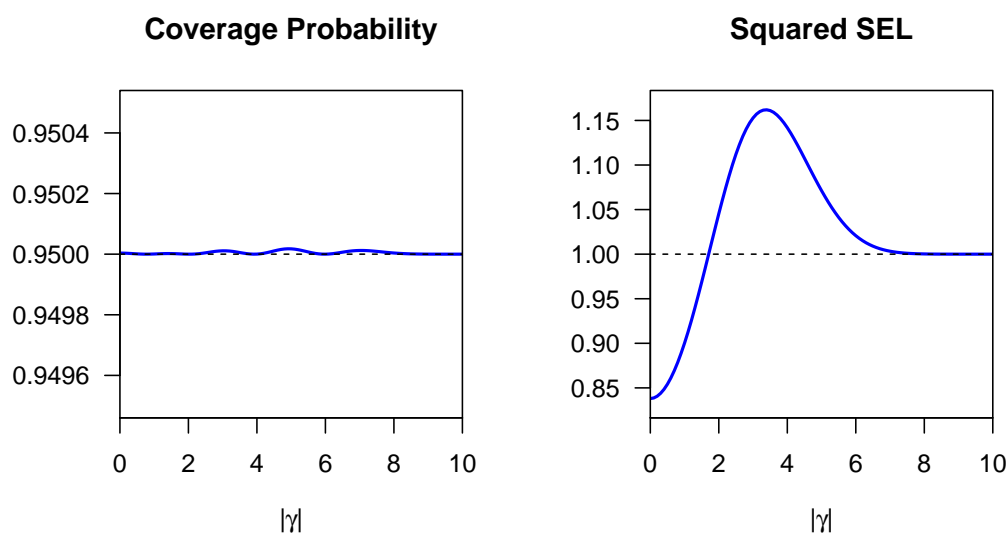**Coverage Probability**   **Squared SEL**

Figure 3: Graphs of the $CP(\gamma; b, s, \rho)$ and the square of $SEL(\gamma; b, s, \rho)$ functions for the CIUUPI, with minimum coverage probability 0.95, where the functions b and s are described by natural cubic spline interpolation, for the factorial experiment example.

Coverage Probability   **Squared SEL**

Figure 4: Graphs of the $CP(\gamma; b, s, \rho)$ and the square of $SEL(\gamma; b, s, \rho)$ functions for the CIUUPI, with minimum coverage probability 0.95, where the functions b and s are described by clamped cubic spline interpolation, for the factorial experiment example.

Figures (5) and (6) show the differences between the Bayesian 95% credible interval and the 95% CIUUPI. Figure (5) shows the graphs of the b and b_B functions (left panel), and the s and s_B functions (right panel), for the factorial experiment example. Note that, similarly to b and s, b_B is an odd continuous function and s_B is an even continuous function. Figure (6) shows the graph of the frequentist coverage probability of the Bayesian 95% credible interval, for the factorial experiment example. This coverage probability is also an even function of γ. Unlike the coverage probability of the CIUUPI, the minimum over γ of the frequentist coverage probability of the Bayesian 95% credible interval is substantially less than 0.95.

Computing the confidence interval (using the observed response)

The observed response for the factorial experiment example data is $y = (87.2, 88.4, 86.7, 89.2)$ and σ is assumed to take the value 0.8. We use the function ciuupi to return the confidence interval (1) for θ that utilizes the uncertain prior information that $\tau = 0$. Continuing from the previous example, this is done in R as follows:

```
# Using the vector (b(1),b(2),...,b(5),s(0),s(1),...,s(5)), compute the CIUUPI
# for this particular data
t <- 0
y <- c(87.2, 88.4, 86.7, 89.2)

ci <- ciuupi(alpha, a, c, x, bsvec, t, y, natural = 1, sig = 0.8); ci
```

We obtain the output

```
           lower    upper
ciuupi  -0.7710755 3.218500
```

For comparison purposes, the function standard_CI will return the standard confidence interval (3) for θ. The code

```
# Compute the standard confidence interval
cistandard(a = a, x = x, y = y, alpha = alpha, sig = 0.8)
```

will return

```
             lower    upper
standard -1.017446 3.417446
```

The 95% confidence interval that utilizes uncertain prior information $[-0.77, 3.22]$ is much shorter than the standard confidence interval $[-1.02, 3.42]$. These are observed values of confidence intervals that have, to an excellent approximation, the same coverage probability. For comparison, a 95% Bayesian credible interval for θ is $[-0.25, 3.51]$. Although this interval is shorter than the CIUUPI, it can be seen from Figure (6) that the minimum over γ of the frequentist coverage of the Bayesian credible interval is substantially less than 0.95.

Discussion

It is very common in applied statistics to carry out preliminary data-based model selection using, for example, hypothesis tests or minimizing a criterion such as the AIC. As pointed out by Leamer (1978, chapter 5), such model selection may be motivated by the desire to utilize uncertain prior information in subsequent statistical inference. He goes even further when he states, on p.123, that "The mining of data that is common among non-experimental scientists constitutes prima facie evidence of the existence of prior information". One may attempt to utilize such prior information by constructing confidence intervals, using the same data, based on the assumption that the selected model had been given to us *a priori*, as the true model. This assumption is false and it can lead to confidence intervals that have minimum coverage probability far below the desired minimum coverage $1 - \alpha$ (see e.g. Kabaila, 2009, Leeb and Pötscher, 2005), making them invalid.

A numerical constrained optimization approach to the construction of valid confidence intervals and sets that utilize uncertain prior information has been applied by Farchione and Kabaila (2008), Kabaila and Giri (2009a), Kabaila and Giri (2013), Kabaila and Giri (2014), Kabaila and Tissera (2014) and Abeysekera and Kabaila (2017). In each case, numerical constrained optimization was performed using programs written in MATLAB, restricting the accessibility of these confidence intervals and sets. The R package **ciuupi** is a first step in making these types of confidence intervals and sets more widely accessible.
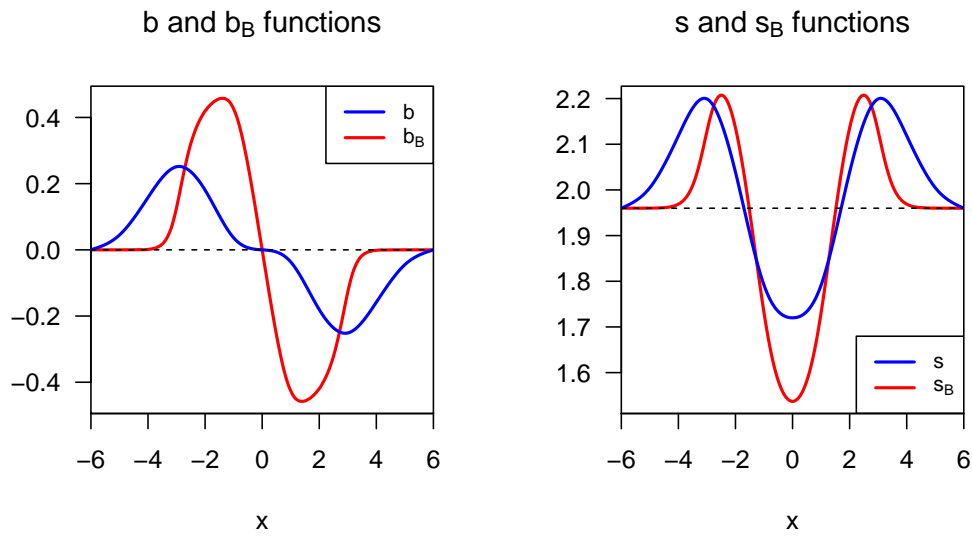
Figure 5: Graphs of the b and b_B functions (left panel), and the s and s_B functions (right panel), for the factorial experiment example.
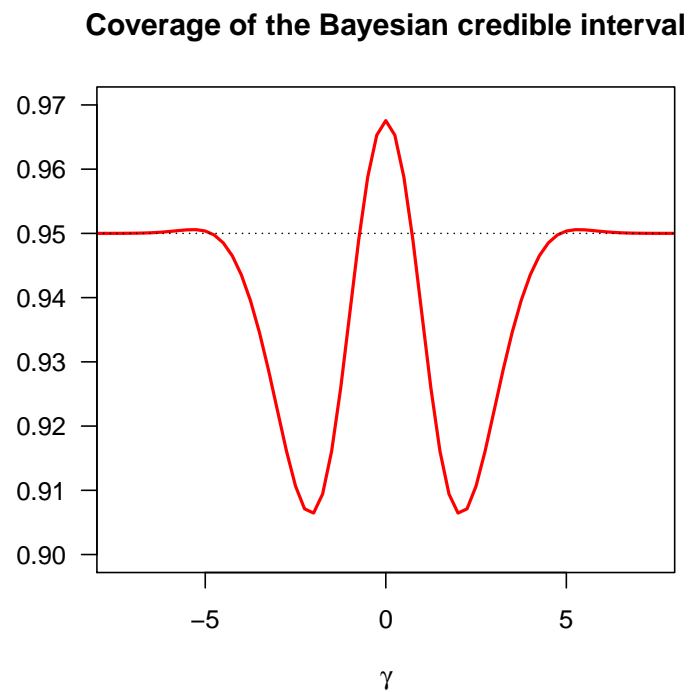

Figure 6: The frequentist coverage probability of the Bayesian 95% confidence interval, for the factorial experiment example.

Bibliography

W. Abeysekera and P. Kabaila. Optimized recentered confidence spheres for the multivariate normal mean. *Electronic Journal of Statistics*, 11:1798–1826, 2017. URL https://doi.org/10.1214/17-EJS1272. [p10]

P. J. Bickel. Parametric robustness: Small biases can be worthwhile. *The Annals of Statistics*, 12:864–879, 1984. ISSN 0090-5364. URL https://doi.org/10.1214/aos/1176346707. [p1]

G. E. P. Box, L. R. Connor, W. R. Cousins, O. L. Davies, F. R. Hinsworth, and G. P. Sillitto. *The Design and Analysis of Industrial Experiments*. Oliver and Boyd, London, 2nd edition, 1963. [p5, 6]

G. Casella and J. T. Hwang. Empirical Bayes Confidence Sets for the Mean of a Multivariate Normal Distribution. *Journal of the American Statistical Association*, 78:688–698, 1983. URL https://doi.org/10.2307/2288139. [p1]

G. Casella and J. T. Hwang. Employing vague prior information in the construction of confidence sets. *Journal of Multivariate Analysis*, 21:79–104, 1987. URL https://doi.org/10.1016/0047-259X(87)90100-X. [p1]

A. Cohen. Improved confidence intervals for the variance of a normal distribution. *Journal of the American Statistical Association*, 67:382–387, 1972. ISSN 0162-1459. URL https://doi.org/10.2307/2284389. [p1]

B. Efron. Minimum volume confidence regions for a multivariate normal mean vector. *Journal of the Royal Statistical Society B*, 68:655–670, 2006. URL https://doi.org/10.1111/j.1467-9868.2006.00560.x. [p1]

D. Farchione and P. Kabaila. Confidence intervals for the normal mean utilizing prior information. *Statistics & Probability Letters*, 78:1094–1100, 2008. URL https://doi.org/10.1016/j.spl.2007.11.003. [p10]

K. Giri. *Confidence Intervals in Regression Utilizing Prior Information*. PhD thesis, Department of Mathematics and Statistics, La Trobe University, 2008. [p3, 4]

C. Goutis and G. Casella. Improved invariant confidence intervals for a normal variance. *The Annals of Statistics*, 19:2015–2031, 1991. URL https://doi.org/10.1214/aos/1176348384. [p1]

J. L. Hodges and E. L. Lehmann. The use of previous experience in reaching statistical decisions. *Annals of Mathematical Statistics*, 23:396–407, 1952. ISSN 0003-4851. URL https://doi.org/10.1214/aoms/1177729384. [p1]

S. G. Johnson. The nlopt nonlinear-optimization package. http://ab-initio.mit.edu/nlopt, 2014. [p4]

P. Kabaila. The coverage properties of confidence regions after model selection. *International Statistical Review*, 77:405–414, 2009. URL https://doi.org/10.1111/j.1751-5823.2009.00089.x. [p10]

P. Kabaila and G. Dharmarathne. A Comparison of Bayesian and Frequentist Interval Estimators in Regression That Utilize Uncertain Prior Information. *Australian & New Zealand Journal of Statistics*, 57(1):99–118, 2015. URL https://doi.org/10.1111/anzs.12104. [p4, 5]

P. Kabaila and K. Giri. Confidence intervals in regression utilizing prior information. *Journal of Statistical Planning and Inference*, 139:3419–3429, 2009a. URL https://doi.org/10.1016/j.jspi.2009.03.018. [p1, 2, 5, 10]

P. Kabaila and K. Giri. Upper bounds on the minimum coverage probability of confidence intervals in regression after model selection. *Australian & New Zealand Journal of Statistics*, 51:271–287, 2009b. URL https://doi.org/10.1111/j.1467-842X.2009.00544.x. [p1]

P. Kabaila and K. Giri. Further properties of frequentist confidence intervals in regression that utilize uncertain prior information. *Australian & New Zealand Journal of Statistics*, pages 259–270, 2013. URL https://doi.org/10.1111/anzs.12038. [p2, 10]

P. Kabaila and K. Giri. Simultaneous confidence intervals for the population cell means, for two-by-two factorial data, that utilize uncertain prior information. *Communications in Statistics – Theory and Methods*, 43:4074–4087, 2014. URL https://doi.org/10.1080/03610926.2012.718846. [p10]

P. Kabaila and D. Tissera. Confidence intervals in regression that utilize uncertain prior information about a vector parameter. *Australian & New Zealand Journal of Statistics*, 56:371–383, 2014. URL https://doi.org/10.1111/anzs.12090. [p10]

P. Kabaila, K. Giri, and H. Leeb. Admissibility of the usual confidence interval in linear regression. *Electronic Journal of Statistics*, 4:300–312, 2010. URL https://doi.org/10.1214/10-EJS563. [p1]

P. J. Kempthorne. Minimax-Bayes Compromise Estimators. *Proc. Bus. and Econ. Statist. Sec. Amer. Statist. Assoc.*, pages 563–573, 1983. [p1]

P. J. Kempthorne. Numerical specification of discrete least favorable prior distributions. *Society for Industrial and Applied Mathematics. Journal on Scientific and Statistical Computing*, 8:171–184, 1987. ISSN 0196-5204. URL https://doi.org/10.1137/0908028. [p1]

P. J. Kempthorne. Controlling risks under different loss functions: The compromise decision problem. *The Annals of Statistics*, 16:1594–1608, 1988. ISSN 0090-5364. URL https://doi.org/10.1214/aos/1176351055. [p1]

E. E. Leamer. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. John Wiley & Sons, 1978. [p10]

H. Leeb and B. M. Pötscher. Model selection and inference: Facts and fiction. *Econometric Theory*, 21: 21–59, 2005. URL https://doi.org/10.1017/S0266466605050036. [p10]

J. W. Pratt. Length of confidence intervals. *Journal of the American Statistical Association*, 56:549–567, 1961. ISSN 0162-1459. URL https://doi.org/10.2307/2282079. [p1]

G. K. Smyth. Numerical integration. *Encyclopedia of Biostatistics*, pages 3088–3095, 2005. URL https://doi.org/10.1002/9781118445112.stat05029. [p4]

C. M. Stein. Confidence sets for the mean of a multivariate normal distribution. *Journal of the Royal Statistical Society. Series B (Methodological).*, 24:365–396, 1962. URL https://doi.org/10.1111/j.2517-6161.1962.tb00458.x. [p1]

Y. L. Tseng and L. D. Brown. Good exact confidence sets for a multivariate normal mean. *The Annals of Statistics*, 5:2228–2258, 1997. URL https://doi.org/10.1214/aos/1069362396. [p1]

Dr. Rheanna Mainzer
School of Mathematics and Statistics
The University of Melbourne
Australia
rheanna.mainzer@unimelb.edu.au

Dr. Paul Kabaila
Department of Mathematics and Statistics
La Trobe University
Australia
P.Kabaila@latrobe.edu.au