Simultaneous adjustment of bias and coverage probabilities for confidence intervals

P. Menéndez, Y. Fan, P. H. Garthwaite and S. A. Sisson

May 3, 2014

Abstract

A new method is proposed for the correction of confidence intervals when the original interval does not have the correct nominal coverage probabilities in the frequentist sense. The proposed method is general and does not require any distributional assumptions. It can be applied to both frequentist and Bayesian inference where interval estimates are desired. We provide theoretical results for the consistency of the proposed estimator, and give two complex examples, on confidence interval correction for composite likelihood estimators and in approximate Bayesian computation (ABC), to demonstrate the wide applicability of the new method. Comparison is made with the double-bootstrap and other methods of improving confidence interval coverage.

Keywords: Confidence interval correction; Coverage probability; Composite likelihood; Approximate Bayesian computation.

1 Introduction

Interval estimates are typically intended to have a specified level of coverage. This is true, for example, of both frequentist confidence intervals and Bayesian credible intervals. However, for many problems the coverage of a confidence or credible interval will only equal its nominal value asymptotically, and coverage can be poor even for quite large samples in some situations. In many complex problems, there can be inherent bias which can be difficult to quantify or calculate. This can arise, for example, in composite likelihood problems (Varin et al. 2011) and approximate Bayesian computation (Sisson and Fan 2011). In this paper we propose a novel procedure for adjusting interval estimates that has wide application and will typically reduce the bias in their coverage.

The procedure assumes that the mechanism that generated the sample data could be simulated if population parameters were known. These parameters are estimated by sample statistics derived from real data, and then pseudo-samples are drawn from the estimated population distribution. From each pseudo-sample a confidence/credible interval is determined for the quantity of interest. The frequentist bias in these intervals is calculated

1 School of Mathematics and Physics, University of Queensland, St Lucia, 4072, AUSTRALIA
2 School of Mathematics and Statistics, University of New South Wales, Sydney, 2052, AUSTRALIA
3 Department of Mathematics and Statistics, Open University, Milton Keynes, MK7 6AA, U.K.
and then used to adjust the interval estimate given by the real data.

The method has similarities to the double-bootstrap (Davison and Hinkley 1997), in which a bias correction is applied to a bootstrap interval by re-sampling from a bootstrap distribution. A difference in our method is that it involves only one level of sample generation, which makes it computationally less demanding, although the saving may dissipate if computationally demanding methods (such as Markov chain Monte Carlo) are used to obtain the interval estimates from sample data.

Several authors have looked at the problem of computing confidence intervals with the correct coverage properties. Much of this work is based on variations of bootstrap procedures. In particular, Hall (1986) considered the bootstrap in terms of Edgeworth expansions, and Beran (1987) provided a method for approximate confidence sets, by using the bootstrap (or asymptotic theory) to estimate the relevant quantiles. This so-called pre-pivoting, based on an estimated bootstrap cumulative distribution function, is iterated to produce improved coverage. Efron (1987) introduced a method to correct the coverage of a bootstrap confidence interval within the bootstrap itself, and Martin (1990) proposed an iterated bootstrap procedure to obtain better bootstrap coverage. The so called double bootstrap, where one or more additional levels of bootstrap are conducted to adjust confidence limits, was discussed by DiCiccio et al. (1992) In a non-bootstrap procedure, Garthwaite and Buckland (1992) proposed a method to compute confidence intervals based on Monte Carlo simulations of the Robbins-Monro search process. In the context of autoregressive models, Hansen (1999) presented a method based on bootstrap replications over a grid to compute confidence intervals in situations where standard bootstrap methods fail. There are also several approaches for computing confidence intervals in the presence of nuisance parameters (Kabaila 1993; Kabaila and Lloyd 2000; Lloyd 2011).

In some of our examples, we apply our procedure to reduce bias in the coverage of Bayesian credible intervals. This may not seem intuitive, since coverage is a frequentist property while a Bayesian interval may reflect personal probabilities. However, there are many situations where posterior distributions should preferably be well calibrated. These include inference with objective or probability matching prior distributions, the verification of Bayesian simulation software (Cook et al. 2006) and techniques and diagnostics in likelihood-free Bayesian inference (Fearnhead and Prangle 2012; Prangle et al. 2012).

In Section 2, we describe the proposed method and give theoretical results related to it, illustrating them through simulated examples and comparing the resulting intervals with bootstrap intervals. In Section 3, we apply our method to two more complex, real analyses. One of these involves estimation with composite likelihoods, which is known to produce confidence intervals that are too narrow, and the other involves approximate Bayesian computation, which typically gives larger posterior credibility intervals than desired. Some concluding comments are given in Section 4.
2 Coverage correction for confidence intervals

Suppose we are interested in estimating an equal-tailed $100(1 - \alpha)\%$ confidence interval for some parameter $\theta \in \Theta \subseteq \mathbb{R}$. Thus for observed data $x$, we seek an estimate $L_c(x)$, such that

$$P(\theta \leq L_c(x)) = \alpha/2,$$

where $L_c(x)$ denotes the lower limit of the interval. Similarly for the upper limit, we seek an estimate $U_c(x)$, such that,

$$P(\theta \geq U_c(x)) = \alpha/2.$$

In the frequentist setting, the parameter $\theta$ is considered a fixed quantity and the expressions above are written in terms of pivotal functions of the data, $x$. In the Bayesian setting, the credible interval is computed from the quantiles of the posterior distribution of $\theta$. In an abuse of notation, we will use the above notations in both frequentist and Bayesian cases.

Suppose that we have a method of obtaining estimates, $L(x)$ and $U(x)$, of the correct lower and upper interval bounds, $L_c(x)$ and $U_c(x)$. We do not assume that these estimates produce the correct coverage probability. However, we do assume that the population parameters, $\theta$, can be well approximated from the data. Our goal is to provide a method that gives adjustments to $L(x)$ and $U(x)$ that improve the interval’s coverage. We first give theoretical results for the proposed methodology, and then give details of its implementation.

2.1 Theoretical results

Assumption 1: We suppose that the observed data $x$ come from the model given by $f(x|\theta)$, $\theta \in \Theta$. For any $\theta \in \Theta$, we assume that it is possible to simulate from $f(\cdot|\theta)$.

Assumption 2: Given $\theta$ and data $x \sim f(x|\theta)$ there exists a consistent estimator $\tilde{\theta}$ of $\theta$.

Assumption 1 requires that we are able to simulate replicate data from the model given the values of the parameters. Assumption 2 requires that we have a good estimator for $\theta$, so that interval estimates obtained using $\tilde{\theta}$ converge to those estimates obtained using the population parameter $\theta$, as the amount of data gets large.

In the following, we only require the lengths of the intervals to be consistent. Consequently, Assumption 2 is not always necessary. For example, this occurs if $\theta$ represents a location parameter whose confidence interval has a length that is independent of $\theta$ (see later example). In the frequentist setting, the maximum likelihood estimator of $\theta$ is consistent and unbiased in many finite sample situations. In the Bayesian setting, the posterior distribution is consistent under mild assumptions, and the posterior mean estimate of $\theta$ is asymptotically unbiased. However, in both cases, finite sample bias in $\tilde{\theta}$ may render our method less accurate.

Theorem 2.1: For some $\theta$ and $x \sim f(x|\theta)$, let $L(x)$ be an estimator of the lower limit of a $100(1 - \alpha)\%$ level confidence interval, and suppose that

$$P\{\theta < L(x)\} \neq \alpha/2.$$
Let $G_W$ denote the distribution function of a random variable $W$. Consider the new estimator

$$L_c(x) = L(x) + \xi_{\alpha/2},$$

where $\xi_{\alpha/2}$ is the $\alpha/2$-th quantile of the distribution function $G_{\{\theta - L(x)\}}$, so that $G_{\{\theta - L(x)\}}(\xi_{\alpha/2}) = \alpha/2$. Then the new estimator, $L_c(x)$, will have the correct coverage probability

$$P\{\theta < L_c(x)\} = \alpha/2.$$

**Proof:** See Appendix.

From the above theorem, it can then be seen that for the estimator of the upper limit of a $100(1 - \alpha)$% confidence interval, $U(x)$, we can write

$$U_c(x) = U(x) + \xi_{1-\alpha/2},$$

where $\xi_{1-\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of the distribution function $G_{\{\theta - U(x)\}}$, so that $G_{\{\theta - U(x)\}}(\xi_{1-\alpha/2}) = 1 - \alpha/2$. In this case, we then have that

$$P\{\theta > U_c(x)\} = \alpha/2.$$

**Theorem 2.2** For some $\theta$ and observed data $x \sim f(x|\theta)$, suppose the lower limit of a $100(1 - \alpha)$% confidence interval $L(x)$ can be obtained, and that this estimate does not necessarily give the correct coverage probability. Suppose that $\hat{\theta} \in \mathcal{R}$ is a consistent estimator of $\theta$, evaluated using the data $x$. Let $y_1, \ldots, y_n$ be $n$ replicate datasets simulated independently from $f(\cdot|\theta)$, and denote the corresponding lower confidence limits by $L_1(y_1), \ldots, L_n(y_n)$, obtained in the same manner as $L(x)$. Define

$$\hat{G}_{\{\hat{\theta} - L(y)\}}(\epsilon) = \frac{1}{n} \sum_{i=1}^{n} I\{\hat{\theta} - L_i(y_i) < \epsilon\}$$

as the empirical distribution of $\hat{\theta} - L(y)$ based on the observed values of $\hat{\theta} - L_i(y_i), i = 1, \ldots, n$. If we define

$$\hat{L}_c(x) = L(x) + \hat{\xi}_{\alpha/2}$$

(3)

where $\hat{\xi}_{\alpha/2} = \hat{G}_{\{\hat{\theta} - L(y)\}}^{-1}(\alpha/2)$, then $\hat{L}_c(x)$ is a consistent estimator of $L_c(x)$, as defined in Equation (1).

**Proof:** See Appendix.

In combination, Theorems 2.1 and 2.2 state that if we simulate data $y_1, \ldots, y_n \sim f(y|\hat{\theta})$ and subsequently obtain the confidence limits $L_1(y_1), \ldots, L_n(y_n)$ in the same way as for the original data $x$, then we can correct the bias in the original lower limit estimate, $L(x)$, by addition of the $\alpha/2$-th sample quantile of $\hat{\theta} - L_1(y_1), \ldots, \hat{\theta} - L_n(y_n)$.

**Corollary 1** Under the assumptions in Theorem 2.2, a central limit theorem holds for $\hat{L}_c(x)$. Specifically, for all $\alpha \in (0, 1)$, $\theta \in \mathcal{R}$ and $x \sim f(x|\theta)$, we have that

$$\sqrt{n}(\hat{L}_c(x) - L_c(x))G'_{\{\theta - L(x)\}}(\xi_\alpha) \longrightarrow N(0, \alpha(1 - \alpha))$$

as $n \rightarrow \infty$, where $G'_{\{\theta - L(x)\}}(\xi) = \frac{\partial}{\partial \xi} G_{\{\theta - L(x)\}}(\xi)$, and $\xi_\alpha$ is the $\alpha$-th quantile of $G_{\{\theta - L(x)\}}$. 


Proof: The result follows immediately from Equation (7) of the proof for Theorem 2.2 (see Appendix).

The above theoretical results provide a simple way of estimating corrections to the lower and upper confidence limits that will produce the correct nominal coverage probability. In addition, these estimators are consistent and asymptotically normal.

2.2 Correction procedure

In summary, the correction algorithm has the following steps:

Step 1 Obtain \( L(x) \) and \( U(x) \), the upper and lower limits of the desired \( 100(1 - \alpha)\% \) confidence interval for the parameter \( \theta \), for an observed dataset \( x \).

Step 2 Evaluate \( \tilde{\theta} \) and generate \( n \) independent datasets \( y_1, \ldots, y_n \sim f(y|\tilde{\theta}) \) from the model.

Step 3 For each dataset \( y_i \), compute the \( 100(1 - \alpha)\% \) lower and upper confidence limits, \( L_i(y_i) \) and \( U_i(y_i) \), for the parameter \( \tilde{\theta} \), using the same method as in Step 1.

Step 4 Set the corrected lower and upper limits to

\[
\begin{align*}
\tilde{L}_c(x) &= L(x) + \hat{G}^{-1}_{\{\tilde{\theta} | L(y)\}}(\alpha/2) \\
\tilde{U}_c(x) &= U(x) + \hat{G}^{-1}_{\{\tilde{\theta} | U(y)\}}(1 - \alpha/2)
\end{align*}
\]

where \( \hat{G}^{-1}_{\{W\}}(\alpha) \) denotes the \( \alpha \)-th sample quantile of the random variable \( W \).

2.3 Simple examples

We illustrate the above procedure with two simple examples. In the first, we consider confidence interval correction for the mean parameter of a normal distribution with known variance. In the second example, it is assumed that the mean is known and that we are interested in the variance parameter.

Example 1: Normal distribution with known variance

Suppose that \( \theta \) is the location parameter of a Normal distribution with unit variance, so that \( x_i \sim N(\theta, 1) \) where \( x = (x_1, \ldots, x_m) \). In this case, the maximum likelihood estimator is \( \tilde{\theta} = \bar{x} = \sum x_i/m \). For illustration, we suppose that the confidence interval we obtain for \( \theta \) does not have the correct coverage, in that we obtain the equivalent confidence interval when data are generated from \( x_i \sim N(\theta, (1 + \epsilon)^2) \) with \( \epsilon \geq 0 \). The value of \( \epsilon \) controls the amount of error in the coverage probability. Following the usual frequentist approach, the \( 100(1 - \alpha)\% \) confidence interval for \( \theta \) is given by \( L(x) = \bar{x} - z_{\alpha/2}(1+\epsilon)/\sqrt{m} \) and \( U(x) = \bar{x} + z_{1-\alpha/2}(1+\epsilon)/\sqrt{m} \), where \( z_{\alpha} \) is the \( \alpha \)-th quantile of the standard normal distribution. Clearly the correction for the interval when \( \epsilon > 0 \) is \( L_c(x) = L(x) + z_{\alpha/2}\epsilon/\sqrt{m} \) and \( U_c(x) = U(x) - z_{1-\alpha/2}\epsilon/\sqrt{m} \).

Figure [I] displays the results of the correction procedure for a 95% confidence interval based on 100 replicate analyses. Each analysis is based on samples of size \( m = 20 \) with
\( \theta = 0 \), so that \( x_1, \ldots, x_m \sim N(0, 1) \), and \( n = 100 \) replicated samples \( y_1, \ldots, y_n \) with elements drawn from \( N(\bar{\theta}, 1) \). Figure 1 (top plots) illustrates the corrected confidence limits \( \hat{L}_c(x) \) and \( \hat{U}_c(x) \) for a range of error term values, \( \epsilon \). Clearly the correction produces an unbiased adjustment, as the boxplots are centred on the true confidence bounds (the horizontal line) in each case. Further, the performance of the method produces qualitatively the same corrected interval limits, irrespective of the value of \( \epsilon \).

The bottom plots display the corrections \( \hat{L}_c(x) \) and \( \hat{U}_c(x) \) with \( \epsilon = 1 \) fixed, for a range of values of \( \bar{\theta} \). For this example, choosing \( \bar{\theta} \) to be any arbitrary value will result in the same quality of unbiased correction. This arises as the distributions of \( \bar{\theta} - L(y) \) and \( \bar{\theta} - U(y) \) do not change with \( \bar{\theta} \), so that the confidence intervals all have the same width as \( \bar{\theta} \) varies. As this is a location parameter only analysis, this is one case where Assumption 2 is not required to produce a consistent adjustment (see Section 2.1).

**Example 2: Normal distribution with known mean**

Suppose now that \( \bar{\theta} \) is the scale (variance) parameter of a Normal distribution with mean zero, so that \( x_i \sim N(0, \theta) \). Here we specify \( \bar{\theta} = \bar{S}^2 = \frac{1}{m-1} \sum_{i=1}^{m}(x_i - \bar{x})^2 \) as the sample variance. In this setting, suppose that the regular confidence limits for \( \bar{\theta} \) are biased downwards by a constant value \( \epsilon > 0 \). Specifically, the \( 100(1-\alpha)/\% \) confidence interval for \( \bar{\theta} \) is given by \( \hat{L}(x) = \frac{(m-1)S^2}{\chi^2_{1-\alpha/2,m-1}} - \epsilon \) and \( \hat{U}(x) = \frac{(m-1)S^2}{\chi^2_{\alpha/2,m-1}} - \epsilon \), where \( \chi^2_{\alpha,k} \) denotes the \( \alpha \)-th percentile of a \( \chi^2_k \) distribution with \( k \) degrees of freedom.

Figure 2 shows the results of the correction procedure for the lower limit of a 95\% confidence interval based on 100 replicate analyses. Each analysis uses samples of size \( m \) with \( \bar{\theta} = 1 \), so that \( x_1, \ldots, x_m \sim N(0, 1) \), and \( n = 2000 \) replicated samples \( y_1, \ldots, y_n \) with elements drawn from \( N(0, \hat{\theta}) \). Figure 2 (left panel) illustrates the corrected lower confidence limit, \( \hat{L}_c(x) \), based on a sample of size \( m = 20 \), for a range of fixed values of \( \hat{\theta} \). The extreme left and right boxplots correspond to the raw biased \( (\hat{L}(x)) \) and true unbiased \( (L_c(x)) \) limits respectively. Clearly, as \( \hat{\theta} \) changes, then so does the location of the adjusted limits. This occurs as, in contrast with the above example, the distributions of \( \bar{\theta} - L(y) \) and \( \bar{\theta} - U(y) \) clearly do change with \( \bar{\theta} \). When \( \bar{\theta} = \theta = 1 \), then the correction procedure produces the correct adjusted limits, as indicated by the rightmost boxplot. Hence, it is necessary to use the right value for \( \bar{\theta} \) when making the correction.

Assumption 2 requires that \( \bar{\theta} \) is a consistent estimator of \( \theta \). Hence we can be sure that \( \bar{\theta} \to \theta \) as \( m \to \infty \), and as a result that the distribution of \( \bar{\theta} - L(y) \) approaches that of \( \theta - L(x) \), so that our correction procedure will perform correctly for large enough \( m \). In practice, the required value of \( m \) can be moderate. Figure 2 (right panel) shows how the correction error, \( \hat{L}_c(x) - L_c(x) \), varies as a function of \( m \). Clearly, the median error is close to zero even for small sample sizes. However, there is some asymmetry for small \( m \), which is also visible in the left panel (e.g. compare the differences in the bias in the boxplots with \( \bar{\theta} = 0.4 \) and \( \bar{\theta} = 1.6 \)), although this is eliminated as \( m \) increases.

Finally, Table 2 compares the empirical coverage probabilities for 95\% confidence intervals for both \( \mu \) and \( \sigma^2 \) in Examples 1 and 2, using our correction procedure and the parametric bootstrap (e.g. Davison and Hinkley 1997).
Figure 1: Lower $\tilde{L}_c(x)$ and upper $\tilde{U}_c(x)$ corrected confidence limit estimates for 100 replicated analyses for the normal location model. Top plots show the corrected limits for $\epsilon = 20, 13.33, 6.67, 0$ with $\tilde{\theta} = \bar{x}$. Bottom plots show the corrected limits for $\tilde{\theta} = -2, -1, 0, 1, 2$ with $\epsilon = 1$. The horizontal lines represent the 0.025-th (left plots) and 0.975-th (right plots) percentiles of a standard normal distribution.
Figure 2: Left panel: Boxplots of the corrected lower confidence limit, $\tilde{L}_c(x)$, when holding $\tilde{\theta}$ fixed at various values $\tilde{\theta} = 0.4, ..., 1.6$ (true value is $\tilde{\theta} = 1$). Leftmost and rightmost boxplots correspond to the biased ($L(x)$) and true unbiased ($L_c(x)$) lower limits respectively. Right panel: Boxplots of the correction error $L_c(x) - L(x)$ as a function of observed data sample size $m$. All boxplots are based on 100 replicate analyses.
Table 1: Empirical coverage probabilities for 95% confidence intervals for the parameters in Examples 1 and 2, for various methods, and for differing values of error, $\epsilon$. Columns denote coverage of (P) the pivot-based intervals, (CP) our correction of the pivot intervals, (B) parametric bootstrap intervals, (CB) our correction of the parametric bootstrap intervals, and (DB) the double bootstrap intervals. Coverage probabilities are based on 1,000 replicate analyses under each method, and $n = 2,000$ generated datasets $y_1, \ldots, y_n$ for the corrected pivotal (CP) and corrected bootstrap (CB). The bootstrap results used 2,000 bootstrap samples (B), 99 bootstrap samples for the corrected bootstrap (CB), and $2000 \times 44$ samples for the double bootstrap (DB). Time (in seconds) indicates the time needed to produce one adjusted replicate interval.

Where there is no error in the construction of the pivot-based confidence intervals (column P) i.e. for $\epsilon = 0$, the corrected intervals (column CP) retain the same correct coverage properties as before the correction. However, we note that as the corrected intervals $(\tilde{L}_c(x), \tilde{U}_c(x))$ are estimated by Monte Carlo, for finite numbers of generated datasets, $y_1, \ldots, y_n$, there will be some non-zero adjustment of each individual confidence interval, even when no systematic error is present. In Table 1 $n = 2,000$ datasets were used for each corrected interval. However, in spite of this random adjustment, the correct coverage is retained over multiple replicates. This point is discussed further in Section 4. The second row of Table 1 shows the same information as the first row, except with a non-zero error, $\epsilon = 6.67$ (for $\mu$) and $\epsilon = 0.20$ (for $\sigma^2$). Clearly the adjusted intervals have the correct nominal coverage.

The third row in Table 1 illustrates the empirical coverage probabilities of 95% confidence intervals based on using the parametric bootstrap (B), our correction of the parametric bootstrap (CB) and the double-bootstrap (DB). Each bootstrap (B) interval was based on 2,000 bootstrap samples, $2,000 \times 44$ samples for the double bootstrap (DB) (following McCullough and Vinod 1998; Booth and Hall 1994), 99 samples for our correction of the bootstrap (CB), and $n = 2000$ datasets, $y_1, \ldots, y_n$, for our correction procedure. These numbers were chosen to provide broadly comparable algorithmic overheads for each method. The bootstrap calculations were implemented using the R package boot, and the double bootstrap confidence intervals were computed as in McCullough and Vinod (1998).

The parametric bootstrap has previously been observed to have lower than nominal coverage (e.g. Schenker 1985; Buckland 1984). In Table 1 this is particularly apparent for $\sigma^2$. Both our correction and the double bootstrap produce improved coverage in each case, although the double bootstrap requires different specification (i.e. the number of bootstrap replicates at each of two levels) than our approach, which alternatively requires the number of auxiliary datasets, $n$, and a consistent estimator of $\theta$. While it is difficult to provide similar algorithmic specifications to permit speed comparisons between the double
bootstrap and our correction procedure, in that they possess different algorithm structures, the recorded times for each method were broadly similar (Table 1, bottom row), with our procedure slightly faster in both current examples. However, the double bootstrap algorithm has a number of optimisations available (e.g. [Nankervis 2005]), whereas our procedure was implemented with unoptimised code. Broadly, the two approaches are comparable in the present analyses.

3 Real examples

We now consider interval estimation in two real, complex modelling situations. The first is an application of composite likelihood techniques in the modelling of spatial extremes. With composite likelihoods, deriving unbiased confidence intervals can require a large amount of algebra, whereas biased intervals that are typically too narrow are easily computable. The second is an application of approximate Bayesian computation (ABC) methods in the modelling of a time series of \(q\)-and-\(k\) distributed observations. In most practical settings the mechanism behind the model fitting process within the ABC framework typically gives posterior credible intervals that are too large.

In both analyses, the parameter \(\theta\) is a vector of \(d > 1\) dimensions. However, our coverage correction procedure is a univariate method as it is based on quantiles. As such, after obtaining the consistent vector estimator \(\hat{\theta}\), we correct the coverage probabilities of each element of the parameter vector in turn, and obtain confidence intervals that achieve the correct nominal marginal coverage probabilities.

3.1 Spatial extremes via composite likelihoods

In the context of analysing spatial extremes, [Padoan et al. 2010] developed a pairwise composite likelihood model, for inference using max-stable stationary processes. Specifically, for \(m\) annual maximum daily rainfall observations, at each of \(K\) spatial locations, the pairwise composite likelihood was specified as

\[
\ell_C(\theta|x) = \sum_{i<j} w_{ij} \log f(x_i, x_j|\theta)
\]

where \(x = (x_1, \ldots, x_K)\) and \(x_i = (x_{i1}, \ldots, x_{im})\), \(f(x_i, x_j|\theta)\) is a known bivariate density function with parameter vector \(\theta\) evaluated at spatial locations \(i\) and \(j\), and \(w_{ij} > 0\) are weights such that \(\sum_{i,j} w_{ij} = 1\). Under the usual regularity conditions, the maximum composite likelihood estimator, \(\hat{\theta}\), can provide asymptotically unbiased and normally distributed parameter estimates when standard likelihood estimators are unavailable (e.g. [Varin et al. 2011]).

Specifically, we have (e.g. [Huber 1967]) that \(\hat{\theta} \sim N(\theta, \hat{I}^{-1}(\hat{\theta}))\), with

\[
I(\theta) = H(\theta)J(\theta)^{-1}H(\theta),
\]

where \(H(\theta)\) and \(J(\theta)\) are respectively the expected information matrix and the covariance matrix of the score vector. In the ordinary maximum likelihood setting, \(H(\theta) = J(\theta)\). In the max-stable process framework, [Padoan et al. 2010] provided an analytic expression for \(J(\theta)\) for a particular (Gaussian) spatial dependence model. Combined with the
standard numerical estimates of $H(\theta)$, this allowed for the construction of standard confidence intervals for $\theta$. However, for composite likelihood techniques in general, obtaining analytic expressions or numerical estimates of $J(\theta)$ can be challenging, whereas estimates of $H(\theta)$ are readily available. In this example, we demonstrate how our proposed method can be employed to correct the too narrow confidence intervals that result from using $I(\theta) = H(\theta)$. We then compare our results with those derived from the known maximum composite likelihood information matrix $I(\theta)$.

We considered four spatial models for stationary max-stable processes that describe different degrees of extremal dependence, with parameter inference based on $m = 100$ observations at each of $K = 50$ randomly generated spatial locations. Each model expresses the degree of extremal dependence via the covariance matrix

$$
\Sigma = \begin{bmatrix}
\sigma_1^2 & \sigma_{12} \\
\sigma_{12} & \sigma_2^2
\end{bmatrix},
$$

where the values for each parameter for each model $M_1, \ldots, M_4$ are given in Table 2.

Model $M_4$ has an additional non-stationary spatial component that is modelled by the extra marginal parameters $\mu, \lambda$ and $\xi$ (corresponding to location, scale and shape parameters) through the response surface

$$
\begin{align*}
\mu &= \alpha_0 + \alpha_1 \cdot \text{lat} + \alpha_2 \cdot \text{lon} \\
\lambda &= \beta_0 + \beta_1 \cdot \text{lon} \\
\xi &= \gamma_0,
\end{align*}
$$

where $\text{lat}$ and $\text{lon}$ denote latitude and longitude coordinates.

| Model   | $\sigma_1^2$ | $\sigma_{12}$ | $\sigma_2^2$ |
|---------|--------------|---------------|--------------|
| $M_1, M_4$ | 9/8         | 0             | 9/8          |
| $M_2$    | 2000         | 1200          | 2800         |
| $M_3$    | 25           | 35            | 14           |

Table 2: Covariance matrix configurations for models $M_1, \ldots, M_4$ for the extremal spatial dependence analysis.

Tables 3 and 4 summarise the empirical coverage probabilities for nominal 95% confidence intervals for models $M_1, \ldots, M_4$ based on 500 replicate analyses. Columns $C_1$ provide the interval coverage using the standard Hessian matrix $I(\theta) = H(\theta)$, and columns $C_2$ provide the same using the composite likelihood information matrix $I(\theta)$ following Padoan et al. (2010). Columns $C_3$ correspond to our correction procedure when applied to the intervals in column $C_1$ using the standard Hessian matrix. For the correction procedure we used $\hat{\theta} = \tilde{\theta}$, the maximum composite likelihood estimate, and $n = 500$ simulated datasets to perform the adjustment.

From Table 3, clearly confidence interval coverage based on the standard Hessian matrix ($C_1$) is too low. The coverage using the sandwich information matrix ($C_2$) is very good, with all the reported values close to 0.95. The coverage values obtained using our adjustment procedure, which is based on the intervals in column $C_1$, are also very close to 0.95.
and mostly closer than with the sandwich information matrix. Similar results are obtained for model $M_4$ in Table 4. Taken together, these results indicate that our adjustment procedure can successfully modify the upper and lower limits of a confidence interval to achieve comparable results to established methods in complex settings. However, it does not make use of the algebraic representation of $J(\theta)$ in this case, and so is more easily extended to alternative models (e.g. where $J(\theta)$ is not available), albeit at a moderate computational cost.

|     | $M_1$       |     | $M_2$       |     | $M_3$       |     |
|-----|-------------|-----|-------------|-----|-------------|-----|
|     | $C_1$ | $C_2$ | $C_3$ |     | $C_1$ | $C_2$ | $C_3$ |     | $C_1$ | $C_2$ | $C_3$ |
| $\sigma_1^2$ | 0.428 | 0.960 | 0.960 |     | 0.098 | 0.947 | 0.950 |     | 0.092 | 0.939 | 0.944 |
| $\sigma_{12}$ | 0.518 | 0.940 | 0.960 |     | 0.122 | 0.955 | 0.956 |     | 0.154 | 0.921 | 0.960 |
| $\sigma_2^2$ | 0.468 | 0.930 | 0.936 |     | 0.092 | 0.955 | 0.938 |     | 0.102 | 0.940 | 0.952 |

Table 3: Empirical coverage probabilities for 95% confidence intervals of the parameters of models $M_1$, $M_2$, and $M_3$ based on 500 replicate analyses. Columns indicate interval confidence estimation methods using: (C_1) the standard Hessian matrix $I(\theta) = H(\theta)$; (C_2) the sandwich information matrix $I(\theta) = H(\theta)J^{-1}(\theta)H(\theta)$; and (C_3) the standard Hessian matrix $I(\theta) = H(\theta)$ followed by our correction procedure.

|     | $C_1$  | $C_2$  | $C_3$  | $C_1$  | $C_2$  | $C_3$  | $C_1$  | $C_2$  | $C_3$  |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $\sigma_1^2$ | 0.372 | 0.544 | 0.388 |     | 0.114 | 0.122 | 0.098 |     | 0.108 | 0.140 | 0.104 |
| $\sigma_{12}$ | 0.930 | 0.925 | 0.945 |     | 0.935 | 0.945 | 0.945 |     | 0.935 | 0.940 | 0.910 |
| $\sigma_2^2$ | 0.924 | 0.924 | 0.958 |     | 0.950 | 0.940 | 0.944 |     | 0.944 | 0.952 | 0.952 |

Table 4: Empirical coverage probabilities for 95% confidence intervals of the parameters of model $M_4$ based on 500 replicate analyses. Columns indicate interval confidence estimation methods using: (C_1) the standard Hessian matrix $I(\theta) = H(\theta)$; (C_2) the sandwich information matrix $I(\theta) = H(\theta)J^{-1}(\theta)H(\theta)$; and (C_3) the standard Hessian matrix $I(\theta) = H(\theta)$ followed by our correction procedure.

### 3.2 Exchange rate analysis using approximate Bayesian computation

Approximate Bayesian computation (ABC) describes a family of methods of approximating a posterior distribution when the likelihood function is computationally intractable, but where sampling from the likelihood is possible (e.g. Beaumont et al. 2002; Sisson and Fan 2011). These methods can be thought of as constructing a conditional density estimate of the posterior (Blum 2010), where the scale parameter, $h > 0$, of the kernel density function controls both the level of accuracy of the approximation, and the computation required to construct it. Lower $h$ results in more accurate posterior approximations, but in return requires considerably more computation. As such, moderate values of the scale parameter are often used in practice. Accordingly, this typically results in oversmoothed estimates of the posterior, and in turn, credible intervals that are too wide.

We consider an analysis of daily exchange rate log returns of the British pound to the Australian dollar between 2005 and 2007. Drovandi and Pettitt (2011) developed an MA(1) type model for these data where the individual log returns were modelled by a $g$-and-$k$ distribution (Rayner and MacGillivray 2002). The $g$-and-$k$ distribution is typically de-
fined through it’s quantile function

\[ Q(z(p); \theta) = a + b \left( 1 + c \frac{1 - \exp(-gz(p))}{1 + \exp(-gz(p))} \right) (1 + z(p)^2)^k z(p), \] (5)

where \( \theta = (a, b, g, k) \) are parameters controlling location, scale, skewness and kurtosis, and \( z(p) \) is the \( p \)-quantile of a standard normal distribution. The parameter \( c = 0.8 \) is typically fixed. We used the sequential Monte Carlo-based ABC algorithm in Drovandi and Pettitt (2011), based on 2,000 particles, to fit the MA(1) model. The data-generation process, used in both ABC and our correction procedure, consists of drawing dependent quantiles \( z_i = (\eta_i + \alpha \eta_{i-1})/\sqrt{1 + \alpha^2} \) for \( i = 1, \ldots, n \), where \( \eta_i \sim N(0, 1) \) for \( i = 0, \ldots, n \), and then substituting \( z(p) = z_i \) in (5).

Table 5 shows the estimated 95% central credible intervals, and their widths, for each model parameter based on the ABC kernel scale parameter \( h = 0.016 \) (following Drovandi and Pettitt 2011) and also the lower value of \( h = 0.009 \). Also shown are the intervals obtained after performing a local-linear, ridge regression-adjustment (Blum et al. 2013; Beaumont et al. 2002) on the posterior obtained with \( h = 0.016 \). The regression-adjustment is a standard ABC technique for improving the precision of an ABC posterior approximation, which aims to estimate the posterior at \( h = 0 \) based on an assumed regression model.

Clearly the parameter credible intervals obtained with \( h = 0.009 \) are narrower than those obtained with \( h = 0.016 \), indicating that the larger intervals indeed have greater than 95% coverage. The regression-adjusted intervals generally have widths somewhere between the intervals constructed with \( h = 0.016 \) and \( h = 0.009 \). The suggestion from Table 5 is that even narrower (i.e. more accurate) credible intervals may result if it were possible to reduce \( h \) further.

Table 6 shows the corrected 95% central credible interval estimates, obtained from the ABC posterior approximations with kernel scale parameter \( h = 0.016, 0.02 \) and 0.03. The correction was based on \( n = 500 \) simulated datasets and using the posterior mean as the estimate \( \hat{\theta} \) of \( \theta \). The results of the correction across the three kernel scale parameter values are similar, suggesting potential computational savings in the ABC posterior simulation stage, as one may perform the analysis with larger values of \( h \). All parameters achieve equivalent or improved precision compared to the most precise ABC posterior estimate obtained with \( h = 0.009 \).

While for a standard Bayesian analysis, the posterior mean is a consistent estimator of \( \theta \), this may not be true in the case of the ABC approximate posterior for \( h > 0 \), as the location and shape of the ABC posterior can change with \( h \). However, the posterior mean is a consistent estimator for \( \theta \) for \( h = 0 \). As such, some care may be needed when specifying \( \theta \) as the posterior mean in the ABC setting. In the current analysis, a preliminary investigation suggested that estimates of the posterior mean stabilised below \( h = 0.03 \), which suggest that the posterior mean is approximately consistent for \( h < 0.03 \). While this determination is slightly ad-hoc, it is practically viable, and an intuitively sensible way of determining whether the computed posterior mean is a consistent estimator of \( \theta \). As such, we are confident that the posterior mean produces an effectively consistent estimate \( \hat{\theta} \) of \( \theta \) in this case.
Table 5: 95% central credible intervals and corresponding interval widths from the $g$-and-$k$ distribution MA(1) model. Results obtained using ABC posterior approximation with kernel scale parameter $h = 0.016$ and $h = 0.009$, and following a ridge regression-adjustment based on an ABC posterior approximation with $h = 0.016$.

|     | $h = 0.016$ Width | $h = 0.009$ Width | Reg. Adj. ($h = 0.016$) Width |
|-----|-------------------|-------------------|-----------------------------|
| a   | (-0.0006, 0.0002) | 0.0008            | (-0.0006, 0.0002) 0.0008     |
| b   | ( 0.0018, 0.0028) | 0.0010            | ( 0.0018, 0.0027) 0.0009     |
| g   | (-0.0267, 0.2573) | 0.2840            | (-0.0286, 0.2505) 0.2791     |
| k   | ( 0.2024, 0.5061) | 0.3037            | ( 0.2148, 0.5092) 0.2944     |
| $\alpha$ | ( 0.1413, 0.2713) | 0.1300            | ( 0.1491, 0.2742) 0.1253     |

Table 6: Adjusted 95% central credibility intervals and corresponding interval widths from the $g$-and-$k$ distribution MA(1) model. Adjusted intervals based on correcting ABC posterior approximations with kernel scale parameter $h = 0.016$, $0.02$ and $0.03$.

|     | $h = 0.016$ Width | $h = 0.02$ Width | $h = 0.03$ Width |
|-----|-------------------|------------------|------------------|
| a   | (-0.0003, 0.0000) | 0.0003           | (-0.0004, -0.0001) 0.0005 |
| b   | ( 0.0020, 0.0024) | 0.0004           | ( 0.0021, 0.0024) 0.0003   |
| g   | ( 0.0303, 0.2156) | 0.1853           | ( 0.0173, 0.1818) 0.1645   |
| k   | ( 0.2769, 0.4099) | 0.1330           | ( 0.2909, 0.4235) 0.1326   |
| $\alpha$ | ( 0.1430, 0.2659) | 0.1229           | ( 0.1335, 0.2708) 0.1373   |

4 Discussion

In this article we have introduced a method of adjusting confidence interval estimates to have a correct nominal coverage probability. This method was developed in the frequentist framework, but may be equally applied to ensure that Bayesian credible intervals possess the (frequentist) coverage property. Our approach is general and makes minimal assumptions: namely that it is possible to generate data under the same procedure (model) that produced the observed data, and that a consistent estimator is available for the parameter of interest. The correction is asymptotically unbiased, although it can work well for moderate sample sizes ($m$), and there is a central limit theorem for the corrected interval limits in terms of the number ($n$) of auxiliary samples used to implement the correction.

As the correction is estimated by Monte Carlo, when there is no bias present, so that $L(x) = L_c(x)$, for finite numbers of replicate datasets $y_1, \ldots, y_n$, finite sample estimates of $\xi_{\alpha/2}$ may be non-zero. This will result in small, non-zero adjustments to intervals that already have the correct nominal coverage. In practice, for moderate $n$, this is likely to have negligible effect (e.g. see the results in Table[1]). However, in this and other settings where there is low bias, the central limit theorem of Corollary 1 describes the precision of the finite sample adjustment as a function of $n$, thereby providing a guide as to when the Monte Carlo variability of $L_c(x)$ will be an improvement over the bias of $L(x)$.

As constructed in Theorems 2.1 and 2.2, our proposed correction is for univariate param-
eters, \( \theta \), as it is based on quantiles. For multivariate \( \theta \), from the perspective of adjusting any given margin, the impact of the remaining (nuisance) parameters is controlled through the estimate \( \hat{\theta} \) of \( \theta \). Asymptotically, the consistency of \( \hat{\theta} \) means that \( \hat{\theta} \to \theta \) as the sample size \( m \to \infty \), from which Theorems 2.1 and 2.2 will then hold for the margin of interest. However, sub-asymptotically this is not the case, and the performance of the adjustment of any margin will depend on the quality of the estimate of \( \theta \) (this is also true in the univariate setting). The results of our analyses in Sections 2.3 and 3 suggest that the procedure can work well, even for moderate \( m \).

In practice, in the examples that we have considered, we have found that our method can produce confidence intervals which perform comparably to existing gold standard approaches – though with greater scope for extension to more complicated models – and provide a reliable method of adjusting approximately obtained credible intervals in challenging settings.

One potential criticism of our approach is that it requires the construction of a large number (\( n \)) of confidence or credible intervals in order to correct one interval. In the case where constructing a single interval is computationally expensive, implementing the correction procedure in full can result in a large amount of computation. This was the case in our exchange rate data analysis using ABC methods, where using an alternative ABC algorithm such as regression-adjustment (based on a single large number of model simulations) would have been more efficient. However, regression-adjustment can itself perform poorly if the assumed regression model is incorrect, while as our correction procedure makes minimal assumptions, we may still have good confidence in the resulting adjusted intervals it provides.

**Acknowledgements**

We are grateful to Chris Drovandi and Tony Pettitt for kindly allowing us to use their Matlab codes for the ABC analysis. Financial support for this research was provided by the Australian Research Council Discovery Project scheme (DP1092805) and the University of New South Wales.

**Appendix: Proofs**

**Proof of Theorem 2.1**

Writing \( L_c(x) = L(x) + \delta \), then

\[
P(\theta < L_c(x)) = P(\theta < L(x) + \delta) = P(\theta - L(x) < \delta).
\]

Hence, by definition, \( P(\theta < L_c(x)) = \alpha/2 \) if \( \delta = \xi_{\alpha/2} \), where \( \xi_{\alpha/2} \) is the \( \alpha/2 \)-th quantile of the distribution of \( \theta - L(x) \).
Proof of Theorem 2.2

Let $G_{\theta - L(x)}$ be the distribution function of $\theta - L(x)$, which has positive first derivatives so that $G'_{\theta - L(x)}(\nu) = \frac{d}{d\nu} G_{\theta - L(x)}(\nu) > 0$ for all $\nu \in \mathbb{R}$. Also let $\hat{G}_{\{\hat{\theta} - L(y)\}}$ be the empirical distribution of $\hat{\theta} - L(y)$ based on the samples $\hat{\theta} - L_i(y_i)$, $i = 1, \ldots, n$. From Theorem 2.1 we have that $L_c(\nu) = L(x) + \xi_{\alpha/2}$ for some $\alpha \in (0, 1)$, where $G_{\theta - L(x)}(\xi_{\alpha/2}) = \alpha/2$. Let $\hat{\xi}_{\alpha/2} = \hat{G}^{-1}_{\{\hat{\theta} - L(y)\}}(\alpha/2)$ be the empirical estimate of $\xi_{\alpha/2}$.

If we define $\tilde{L}_c(\nu) = L(x) + \hat{\xi}_{\alpha/2}$, then for any $\omega \in \mathbb{R}$, we have

$$\Pr\left( \sqrt{n}(\tilde{L}_c(x) - L_c(x)) \leq \omega \right) = \Pr\left( \sqrt{n}(\hat{\xi}_{\alpha/2} - \xi_{\alpha/2}) \leq \omega \right) = \Pr(\hat{\xi}_{\alpha/2} \leq \xi_{\alpha/2} + \omega/\sqrt{n}) = \Pr\left( G_{\theta - L(x)}(\hat{\xi}_{\alpha/2}) \leq G_{\theta - L(x)}(\xi_{\alpha/2} + \omega/\sqrt{n}) \right) = \Pr\left( G_{\theta - L(x)}(\hat{L}_c(\nu)) \leq \alpha/2 + \omega G'_{\theta - L(x)}(\xi_{\alpha/2}) + o(1)/\sqrt{n} \right)$$

where the last equality follows from a first order Taylor expansion of $G$ at $\xi_{\alpha/2}$.

If $Y$ represents the number of times that $G(\hat{\xi}_{\alpha/2})$ is smaller than $\xi = \alpha/2 + \omega G'_{\theta - L(x)}(\xi_{\alpha/2}) + o(1)/\sqrt{n}$, then since $G(\xi) \sim U(0, 1)$, we have $Y \sim \text{Binomial}(n, \xi)$. Hence

$$\frac{Y - n\xi}{\sqrt{n\xi(1 - \xi)}} \rightarrow N(0, 1) \quad (6)$$

in distribution as $n \rightarrow \infty$ (der Vaart 2000).

Let $r_n$ be the integer rank of the $\alpha/2$-th quantile from a data set $X = \{X_1, \ldots, X_n\}$ of length $n$, such that $\hat{\xi}_{\alpha/2} = X_{(r_n)}$. If we assume that $\frac{n\alpha/2 - r_n}{\sqrt{n}} \rightarrow 0$ as $n \rightarrow \infty$, then from (6) we have

$$\Pr\left( \sqrt{n}(\hat{\xi}_{\alpha/2} - \xi_{\alpha/2}) \leq \omega \right) = \Pr\left( G_{\theta - L(x)}(\hat{\xi}_{\alpha/2}) \leq \xi \right) = \Pr\left( Y \geq r_n \right) = \Pr\left( \frac{Y - n\xi}{\sqrt{n\xi(1 - \xi)}} \geq \frac{r_n - n\xi}{\sqrt{n\xi(1 - \xi)}} \right) = \Phi\left( \frac{n\alpha/2 - r_n}{\sqrt{n(\alpha/2)(1 - \alpha/2)}} + o_p(1) \right) = \Phi\left( \frac{\omega G'_{\theta - L(x)}(\xi_{\alpha/2})}{\sqrt{(\alpha/2)(1 - \alpha/2)}} + o_p(1) \right)$$

It then follows that the consistency of the estimator $\tilde{L}_c$ can be established as

$$\lim_{n \rightarrow \infty} \Pr\left( \sqrt{n}(\tilde{L}_c(x) - L_c(x)) \geq \epsilon \right) \leq \lim_{n \rightarrow \infty} \frac{E[\sqrt{n}(\tilde{L}_c(x) - L_c(x))]}{\epsilon} = 0$$

by the Markov inequality (Ash and Doleans-Dade 2000).
References

Ash, R. B. and C. Doleans-Dade (2000). *Probability and Measure theory*. Academic Press.

Beaumont, M. A., W. Zhang, and D. J. Balding (2002). Approximate Bayesian computation in population genetics. *Genetics* 162(4), 2025–2035.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* 74(3), 457–468.

Blum, M. G. B. (2010). Approximate Bayesian computation: a nonparametric perspective. *Journal of the American Statistical Association* 105, 1178–1187.

Blum, M. G. B., M. A. Nunes, D. Prangle, and S. A. Sisson (2013). A comparative review of dimension reduction methods in approximate Bayesian computation. *Statistical Science* 28, 189–208.

Booth, J. G. and P. Hall (1994). Monte Carlo approximation and the iterated bootstrap. *Biometrika* 81, 331–340.

Buckland, S. T. (1984). Monte Carlo confidence intervals. *Biometrics* 40, 811–817.

Cook, S., A. Gelman, and D. Rubin (2006). Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics* 15, 675–692.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap methods and their application*. Cambridge University Press.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap methods and their application*. Cambridge University Press.

DiCiccio, T. J., M. A. Martin, and G. A. Young (1992). Fast and accurate approximate double bootstrap confidence intervals. *Biometrika* 79(2), 285–295.

Drovandi, C. C. and A. N. Pettitt (2011). Likelihood-free Bayesian estimation of multivariate quantile distributions. *Computational Statistics & Data Analysis* 55, 2541–2556.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82(397), 171–185.

Fearnhead, P. and D. Prangle (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation (with discussion). *Journal of the Royal Statistical Society: Series B* 74, 419–474.

Garthwaite, P. H. and S. T. Buckland (1992). Generating Monte Carlo confidence intervals by the Robbins-Monro process. *Applied Statistics*, 159–171.

Hall, P. (1986). On the bootstrap and confidence intervals. *The Annals of Statistics*, 1431–1452.

Hansen, B. E. (1999). The grid bootstrap and the autoregressive model. *Review of Economics and Statistics* 81(4), 594–607.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, pp. 221–33.

Kabaila, P. (1993). Some properties of profile bootstrap confidence intervals. *Australian Journal of Statistics* 35(2), 205–214.
Kabaila, P. and C. J. Lloyd (2000). A computable confidence upper limit from discrete data with good coverage properties. *Statistics & probability letters* 47(2), 189–198.

Lloyd, C. (2011). Computing highly accurate confidence limits from discrete data using importance sampling. [http://works.bepress.com/chris_lloyd/23](http://works.bepress.com/chris_lloyd/23).

Martin, M. A. (1990). On bootstrap iteration for coverage correction in confidence intervals. *Journal of the American Statistical Association* 85(412), 1105–1118.

McCullough, B. and H. D. Vinod (1998). Implementing the double bootstrap. *Computational Economics* 12(1), 79–95.

Nankervis, J. C. (2005). Computational algorithms for double bootstrap confidence intervals. *Computational Statistics & Data Analysis* 49, 462–475.

Padoan, S. A., M. Ribatet, and S. A. Sisson (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* 105, 263 – 277.

Prangle, D., M. G. B. Blum, G. Popovic, and S. A. Sisson (2012). Diagnostic tools for approximate Bayesian computation using the coverage property. *Tech Report*, [http://arxiv.org/abs/1301.3166](http://arxiv.org/abs/1301.3166).

Rayner, G. D. and H. L. MacGillivray (2002). Numerical maximum likelihood estimation for the $g$-and-$k$ and generalized $g$-and-$h$ distributions. *Statistics and Computing* 12(1), 57–75.

Schenker, N. (1985). Qualms about bootstrap confidence intervals. *Journal of the American Statistical Association* 390, 360–361.

Sisson, S. A. and Y. Fan (2011). Likelihood-free Markov chain Monte Carlo. In S. P. Brooks, A. Gelman, G. Jones, and X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*, pp. 319–341. Chapman and Hall/CRC Press.

Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica* 21(1), 5–42.