Likelihood and p-value functions in the composite likelihood context

D.A.S. Fraser and N. Reid
Department of Statistical Sciences
University of Toronto

November 19, 2016

Abstract

The need for combining likelihood information arises widely in the analysis of complex models, and also in meta-analysis where information is to be combined from various studies. We work to first order and show that full first order accuracy for combining scalar or vector parameter information is available from likelihood analysis based on score variables.

1 Introduction

Statistical models presented in the form of a family of densities $\{f(y;\theta);\theta \in \Theta \subset \mathbb{R}^p)\}$ are usually analyzed using the likelihood function $L(\theta) \propto f(y;\theta)$, or equivalently the log-likelihood function $\ell(\theta) = \log\{L(\theta)\}$. Evaluated at the observed data this provides all the data-dependent information for a standard Bayesian analysis, and almost all the data-dependent information for a typical frequentist-based analysis where full third-order inference requires the addition of a sample-space derivative.

In some modelling situations however the full density version of the model may not be available. In such cases work-arounds have been developed, using for example marginal models for single coordinates, or marginal models for pairs of coordinates or other variants called pseudo or composite likelihood functions, studied in Lindsay (1988) and reviewed in Varin et al. (2011). For example, if we have a statistical model $f(y_1, \ldots, y_d; \theta)$, the composite pairwise log-likelihood is

$$\ell_{\text{pair}}(\theta) = \sum_{r<s} \log\{f_2(y_r, y_s; \theta)\},$$

where $f_2(y_r, y_s; \theta)$ is the marginal model for the pair of components (y_r, y_s). We can also express this as

$$\ell_{\text{pair}}(\theta) = \sum_{i=1}^{m} \ell_i(\theta),$$

where $m = d(d-1)/2$, and $\ell_i(\theta) = \log\{f_2(y_r, y_s; \theta)\}$, with i indexing the pairs (r, s) so that y_i is short-hand for some pair of components of the vector $y = (y_1, \ldots, y_d)$. If we were considering what is called the independence likelihood we would have $\ell_{\text{ind}}(\theta) = \Sigma_r \log\{f_1(y_r; \theta)\} = \Sigma_i \ell_i(\theta)$ with $\ell_i(\theta) = \log\{f_1(y_i; \theta)\}$ and $m = d$.

Our approach here is to suppose we have available component log-likelihood functions, derived from a sample of size n, and to assume that each component has the usual asymptotic properties, such as being $O_p(n)$ with the typical smoothness asymptotic behaviour of its cumulants. We let θ_0 be some trial or reference value for the parameter and examine the model in moderate deviations about θ_0. We will see then that to first order the model has simple normal form and is independent in structure of the initial reference value θ_0 given that it is in moderate deviations. This leads to a fully first-order accurate combine d log-likelihood. In addition, the approach leads to a simple first-order accurate procedure for combining p-values. The method uses variances and covariances of score variables, as are also needed for the usual inferential adjustments in the composite likelihood approach.

Specifically we expand a component log-likelihood as

$$\ell_i(\theta) \doteq \ell_i(\theta_0) + (\theta - \theta_0)^{\mathrm{T}} \ell_i'(\theta_0) + \frac{1}{2}(\theta - \theta_0)^{\mathrm{T}} \ell_i''(\theta_0)(\theta - \theta_0).$$

and work to first order so that $\theta - \theta_0 = O(n^{-1/2})$; in this case the first term on the right side above is $O_p(n)$, the second two terms are $O_p(1)$, and the neglected terms are $O_p(n^{-1/2})$. In addition, to this order, we can replace the observed second derivative matrix with its expected value. Thus each component log-likelihood depends on the data only through the $p \times 1$ score vector

$$s_i = \frac{1}{\sqrt{n}} \left. \frac{\partial \ell_i(\theta; y_i)}{\partial \theta} \right|_{\theta=\theta_0},$$

standardized here to be $O_p(1)$.

The Bartlett identities for the component models give

$$E(s_i; \theta_0) = 0, \ \operatorname{var}(s_i; \theta_0) = - \left. E\{ \frac{\partial^2 \ell(\theta; y_i)}{\partial \theta \partial \theta^{\mathrm{T}}}; \theta \} \right|_{\theta_0} = v_{ii}, \tag{1}$$

where v_{ii} is the $p \times p$ expected Fisher information matrix, from the ith component log-likelihood, standardized to be $O(1)$. This standardization and our assumed continuity also give (Cox & Hinkley, 1974, Ch. 4)

$$E(s_i; \theta) = v_{ii}(\theta - \theta_0) + O(n^{-1/2}), \quad \operatorname{var}(s_i; \theta) = v_{ii} + O(n^{-1/2}). \tag{2}$$

We now stack the score vectors $s = (s_1, \ldots, s_m)^{\mathrm{T}}$ and write

$$E(s; \theta) = V(\theta - \theta_0) + O(n^{-1/2}), \quad \operatorname{var}(s; \theta) = W + O(n^{-1/2}), \tag{3}$$

where $V = (v_{11}, \ldots, v_{mm})^{\mathrm{T}}$ is the $mp \times p$ matrix of the stacked v_{ii}, and W is the $mp \times mp$ matrix with v_{ii} on the diagonal, and off-diagonal matrix elements $v_{ij} = \operatorname{cov}(s_i, s_j)$. This first order version of the model is structurally free of the reference value θ_0.

2 Full first-order log-likelihood function

We now analyze the statistical model in its full first order accurate form, and for temporary notational convenience write just θ for the departure $\theta - \theta_0$ from the trial value. From the preceding summary we can then write the model for s as a normal theory regression model,

$$s = V\theta + e$$

where $e \sim N(0, W)$. From this full first-order approximation we have directly a first-order log-likelihood function

$$\ell^*(\theta) \;=\; c - \frac{1}{2}(s - V\theta)^{\mathrm{T}} W^{-1}(s - V\theta), \tag{4}$$

$$\;=\; c - \frac{1}{2}\theta^{\mathrm{T}} V^{\mathrm{T}} W^{-1} V\theta + \theta^{\mathrm{T}} V^{\mathrm{T}} W^{-1} s. \tag{5}$$

The score function from (5) is $s^*(\theta) = V^T W^{-1}(s - V\theta)$ with mean 0 and variance $V^T W^{-1} V$; and the maximum likelihood value is

$$\hat{\theta}^* = (V^{\mathrm{T}} W^{-1} V)^{-1} V^{\mathrm{T}} W^{-1} s,$$

with mean θ and variance $(V^{\mathrm{T}} W^{-1} V)^{-1} = \bar{W}$.

,

The new log-likelihood function can also be rewritten in the equivalent form

$$\ell^*(\theta) \;=\; c - \frac{1}{2}(\theta - \hat{\theta}^*)^{\mathrm{T}} V^{\mathrm{T}} W^{-1} V(\theta - \hat{\theta}^*) = c - \frac{1}{2}(\theta - \hat{\theta}^*)^{\mathrm{T}} \overline{W}^{-1}(\theta - \hat{\theta}^*); \tag{6}$$

this makes the location form of the log-likelihood more transparent.

If θ is scalar, then

$$\ell^*(\theta) \;=\; c - \frac{1}{2} V^{\mathrm{T}} W^{-1} V\theta^2 + V^{\mathrm{T}} W^{-1} s\theta, \tag{7}$$

and the individual components are $\ell_i(\theta) = c_i - (1/2)(s_i - v_{ii}\theta)^2 v_{ii}^{-1}$ which can be rearranged giving $s\theta = \underline{\ell}(\theta) + (1/2)V\theta^2 + c$ where $\underline{\ell}(\theta) = \{\ell_1(\theta), \ldots, \ell_m(\theta)\}^{\mathrm{T}}$ is the vector of component log-likelihood functions. Then substituting for $s\theta$ in (7) gives an expression for $\ell^*(\theta)$ as a weighted combination of component log-likelihood functions,

$$\ell^*(\theta) = V^{\mathrm{T}} W^{-1} \underline{\ell}(\theta), \tag{8}$$

with related score function $s^*(\theta) = V^T W^{-1}(s - V\theta)$. The original component log-likelihoods, rather than their Taylor series approximations, can be used here to the same order, and thus (8) provides an optimally weighted combination of component log-likelihoods. While this makes (8) somewhat attractive compared with (7), and agrees with it up to quadratic terms, there are implicit cubic terms in (8) so inferences from the two versions will be different in finite samples, though equivalent to the order to which we are working.

The linear combination of (8) is not generally available for vector parameters; different combinations of components are needed for different coordinates of the parameter, as indicated by different rows in the matrix V^{T} in (7).

Lindsay (1988) studied the choice of weights in scalar composite likelihood by seeking an optimally weighted combination of score functions $\partial \ell_i(\theta)/\partial\theta$, in his notation $S_i(\theta)$; the optimal weights depend on θ. Our approach is to work within $n^{-1/2}$-neighbourhoods of a reference parameter value, and obtain the first-order model for the observed variables s_i, thus leading directly to the full first-order log-likelihood. The use of a quadratic form in a set of estimating equations is used in indirect inference in econometrics, and is called an indirect likelihood in Jiang and Turnbull (2004).

For a scalar or vector parameter of interest ψ of dimension r, with nuisance parameter λ so that $\theta = (\psi^{\mathrm{T}}, \lambda^{\mathrm{T}})^{\mathrm{T}}$, we can then see directly that the first-order log-likelihood function for the component ψ is just

$$\ell^*(\psi) = c - \frac{1}{2}(\psi - \hat{\psi}^*)^{\mathrm{T}} \overline{W}^{\psi\psi} (\psi - \hat{\psi}^*) \tag{9}$$

where $\overline{W}^{\psi\psi}$ is the $\psi\psi$ component of the information $\bar{W}^{-1}$ and the maximum log-likelihood $\hat{\psi} = \psi(\hat{\theta})$ is obtained from the maximizing value $\hat{\theta}$ for the full log-likelihood at (5). For the scalar component, Pace et al. (2016) consider the use of profile log-likelihood components.

3 Illustrations

The first illustrations use latent independent normal variables with mean θ and variance 1, as this captures the essential elements and makes clear the role of correlation in the re-weighting: the basic underlying densities are assumed to be independent responses x from a $N(\theta, 1)$ distribution, with corresponding log-likelihood $-\theta^2/2 + \theta x$.

Example 1: Independent components. Consider component variables $y_1 = a_1 x_1$ and $y_2 = a_2 x_2$. Then $m = 2$ and the component log-likelihoods are $\ell_i(\theta; y_i) = -(y_i - a_i\theta)^2/2a_i^2 = -\theta^2/2 + y_i\theta/a_i$ giving $s_i = y_i/a_i$, $V = (1, 1)^{\mathrm{T}}$, and $W = \mathrm{diag}(1, 1)$, which lead to $V^{\mathrm{T}}W^{-1} = (1, 1)$. Thus $\ell^*(\theta) = \ell_1(\theta) + \ell_2(\theta)$ is the independence log-likelihood, as would be expected.

Example 2: Dependent and exchangeable components. Consider $y_1 = x_1 + x_2$ and $y_2 = x_1 + x_3$. The component log-likelihood functions are $\ell_i(\theta) = -\theta^2 + \theta y_i$ giving $s_i = y_i$,

$$V = (2, 2)^{\mathrm{T}}, \quad W = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \text{ and } \quad V^{\mathrm{T}}W^{-1} = (2/3, 2/3). \tag{10}$$

These lead to the combined first-order log-likelihood function

$$\ell^*(\theta) = (2/3, 2/3)^{\mathrm{T}}\underline{\ell}(\theta) = -\frac{4}{3}\theta^2 + \frac{2}{3}\theta(y_1 + y_2). \tag{11}$$

In contrast the unadjusted composite log-likelihood obtained by adding the marginal log-likelihoods is

$$\ell_{UCL}(\theta) = -2\theta^2 + \theta(y_1 + y_2),$$

with score variable $y_1 + y_2$, which has variance 6, but second derivative 4: the second Bartlett identity does not hold and $\ell_{UCL}(\theta)$ is not a proper log-likelihood. The usual method to adjust this is to rescale the unadjusted log-likelihood $\ell_{ACL} = a\ell_{UCL}$ giving $-2a\theta^2 + \theta a(y_1 + y_2)$ with negative second derivative $4a^2$ and score variance $6a^2$. These latter become equal with $a = 2/3$, which gives the adjusted composite log-likelihood

$$\ell_{ACL}(\theta) = \frac{2}{3}\ell_{UCL}(\theta) = -\frac{4}{3}\theta^2 + \frac{2}{3}\theta(y_1 + y_2),$$

which agrees with $\ell^*(\theta)$. The next example shows that this agreement does not hold generally.

Example 3: Dependent but not exchangeable components. Now let $y_1 = x_1$ and $y_2 = x_1 + x_3$. The individual log-likelihoods are $\ell_1(\theta) = -\theta^2/2 + \theta y_1$ and $\ell_2(\theta) = -\theta^2 + \theta y_2$ with $s_1 = y_1$ and $s_2 = y_2$. We then have $V^{\mathrm{T}} = (1, 2)$, the off-diagonal elements of W equal to 1, and $V^{\mathrm{T}}W^{-1} = (0, 1)$, leading to

$$\ell^*(\theta) = -\theta^2 + \theta y_2, \tag{12}$$

with maximum likelihood estimate $\hat{\theta} = y_2/2$, which has variance 1/2. The new likelihood shows reflects that $y_2 = x_1 + x_3$ provides full first-order information concerning θ from the effective data y_2.

In contrast, the unadjusted composite likelihood is $\ell_{UCL}(\theta) = -(3/2)\theta^2 + \theta(y_1 + y_2)$, with associated maximum likelihood estimate $\hat{\theta}_{UCL} = (y_1 + y_2)/3$. The rescaling factor to adjust for the Bartlett property is $a = 3/5$ giving the adjusted composite likelihood

$$\ell_{ACL}(\theta) = \frac{3}{5}\ell_{UCL}(\theta) = -\frac{9}{10}\theta^2 + \frac{3}{5}\theta(y_1 + y_2),$$

which is again maximized at $(y_1 + y_2)/3$. Although the second Bartlett identity has been implemented, the adjusted composite likelihood leads to the same inefficient estimate of θ as the unadjusted version, while the new log-likelihood $\ell^*(\theta)$ provides the proper information 2 and estimate $y_2/2$. Some discussion related to this point is given in Freedman (2006).

An asymptotic version of these two illustrations is obtained by having n replications of y_1 and y_2, or equivalently assuming x_1, x_2 and x_3 have variances $1/n$ instead of 1.

Example 4: Bivariate Normal. Suppose we have n pairs (y_{i1}, y_{i2}) independently distributed as bivariate normal with mean vector (θ, θ) and a known covariance matrix. The sufficient statistic is $(\bar{y}_{.1}, \bar{y}_{.2})$, and the component log-likelihood functions are taken as those from the marginal densities of $\bar{y}_{.1}$ and $\bar{y}_{.2}$, so that $\ell_1(\theta) = -n(\bar{y}_{.1} - \theta)^2/(2\sigma_1^2)$ and $\ell_2(\theta) = -n(\bar{y}_{.2} - \theta)^2/(2\sigma_2^2)$. The score components s_1 and s_2 are, respectively, $n(\bar{y}_{.1} - \theta)/\sigma_1^2$ and $n(\bar{y}_{.2} - \theta)/\sigma_2^2$, with variance-covariance matrix

$$W = n \begin{pmatrix} 1/\sigma_1^2 & \rho/(\sigma_1\sigma_2) \\ \rho/(\sigma_1\sigma_2) & 1/\sigma_2^2 \end{pmatrix}, \tag{13}$$

giving

$$V^{\mathrm{T}} W^{-1} = (1 - \rho^2)^{-1}(1 - \rho\sigma_1/\sigma_2, 1 - \rho\sigma_2/\sigma_1),$$

leading to

$$\ell^*(\theta) = -\frac{n}{2(1-\rho^2)}\left\{\left(\frac{\bar{y}_{.1} - \theta}{\sigma_1}\right)^2 (1 - \rho\frac{\sigma_1}{\sigma_2}) + \left(\frac{\bar{y}_{.2} - \theta}{\sigma_2}\right)^2 (1 - \rho\frac{\sigma_2}{\sigma_1})\right\}. \tag{14}$$

As a function of θ this can be shown to be equivalent to the full log-likelihood based on the bivariate normal distribution of $(\bar{y}_{.1}, \bar{y}_{.2})$, and leads to an estimate of θ that is a weighted combination of $\bar{y}_{.1}$ and $\bar{y}_{.2}$.

Example 5: Two parameters. Suppose that x_i follow a $N(\theta_1, 1)$ distribution, and independently z_i follow a $N(\theta_2, 1)$ distribution. We base our component log-likelihoods on the densities of the vectors

$$y_1 = \begin{pmatrix} x_1 \\ z_2 + z_3 \end{pmatrix}, \quad y_2 = \begin{pmatrix} x_1 + x_3 \\ z_2 \end{pmatrix},$$

giving the score variables $s_1 = (y_{11}, y_{12})^T$ and $s_2 = (y_{21}, y_{22})^T$. The needed variances and covariances are:

$$V = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad W = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad W^{-1} = \begin{pmatrix} 2 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 2 \end{pmatrix},$$

and then for (6) we have $V^T W^{-1} V = \text{diag}(2, 2)$. This gives the new log-likelihood

$$\ell^*(\theta_1, \theta_2) = -(\theta_1 - \hat{\theta}_1^*)^2 - (\theta_2 - \hat{\theta}_2^*)^2,$$

which represents the log-likelihood from s_{12} plus the log-likelihood from s_{21} as we might reasonably have expected from the presentations in terms of the latent x_i and z_i variables. Meanwhile, the usual composite log-likelihood derived from the sum of those for s_1 and s_2 has extra terms.

Example 6. Two parameters, without symmetry. Within the structure of the previous example, suppose our component vectors are

$$y_1 = \begin{pmatrix} x_1 \\ z_1 \end{pmatrix}, \quad y_2 = \begin{pmatrix} x_1 + x_2 \\ z_2 \end{pmatrix},$$

where the first coordinates depend on θ_1 and the second coordinates depend on θ_2; the corresponding score variables are $s_1 = (y_{11}, y_{12})^T$ and $s_2 = (y_{21}, y_{22})^T$. The variances and covariances again are directly available, and we have

$$V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad W = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad W^{-1} = \begin{pmatrix} 2 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

From Example 3 we can see that for inference about θ_1 the weights are $(0,1)$ and from Example 1 the weights for θ_2 are $(1,1)$; clearly these are incompatible.

For the direct approach using (7) we have $V^TW^{-1}V = \text{diag}(2,2)$, giving

$$\ell^*(\theta_1,\theta_2) = -(\theta_1 - \hat{\theta}_1^*)^2 - (\theta_2 - \hat{\theta}_2^*)^2$$

This simple sum of component likelihoods for (θ_1,θ_2) is to be expected as the measurements concerning θ_1 are independent of those for θ_2; in addition, all the information concerning θ_1 comes from $x_1 + x_2$ as the first coordinate of y_2, and all for θ_2 comes from z_1 and z_2 in the second coordinates of y_1 and y_2.

4 Comparison to composite likelihood

Composite likelihood combines information from different components, often by adding the log-likelihood functions, but care is needed in constructing inference from the resulting function, as the curvature at the maximum does not give an accurate reflection of the precision. Corrections for this in the scalar parameter setting involve either rescaling the composite log-likelihood function, or accommodating the dependence among the components in the estimate of the variance of the composite likelihood estimator. In the vector parameter setting adjustments to the composite log-likelihood function are more complex than a simple rescaling; see Pace et al. (2011).

This rescaling is not enough: the location of the composite log-likelihood function is incorrect to first order, and confidence intervals so obtained are not correctly located to first order. This is corrected by the the use of $\ell^*(\theta)$ from Section 2.

As we are using only first-order log-likelihood functions, it suffices to illustrate this with normal distributions. Suppose $y^{\mathrm{T}} = (y_1, ..., y_m)$, where the marginal models for the individual coordinates y_i are normal with mean θv_{ii} and variance v_{ii}, and $\text{cov}(y_i, y_j) = v_{ij}$, all elements of the matrix W. The unadjusted composite log-likelihood function is

$$\ell_{UCL}(\theta) = -\frac{1}{2}\theta^2\Sigma_{i=1}^m v_{ii} + \Sigma_{i=1}^m y_i\theta.$$

with maximum likelihood estimate $\hat{\theta}_{CL} = \Sigma y_i/\Sigma v_{ii}$ and curvature at the maximum point Σv_{ii}: this curvature is not the inverse variance of $\hat{\theta}_{CL}$ as the second Bartlett identity does not hold.

As indicated in Example 2, the rescaled version that recovers the second Bartlett identity is

$$\ell_{ACL}(\theta) = \frac{H}{J}\ell_{UCL}(\theta) = -\frac{1}{2}\theta^2\frac{(\Sigma v_{ii})^2}{\Sigma v_{ij}} + \theta\sum y_i\frac{\Sigma v_{ii}}{\Sigma v_{ij}},$$

where $H = E\{-\ell''_{UCL}(\theta)\}$ and $J = \mathrm{var}\{\ell'_{UCL}(\theta)\}$; in this context neither H nor J depends on θ. The maximum likelihood estimate from this function is the same, $\hat{\theta}_{UCL}$, but the inverse of the second derivative gives the correct asymptotic variance.

What is less apparent is that the location of the log-likelihood function needs a correction. This is achieved using the new weighted version from Section 2:

$$\ell^*(\theta) = -\frac{1}{2}\theta^2(V^{\mathrm{T}}W^{-1}V) + \theta V^{\mathrm{T}}W^{-1}y,$$

which has maximum likelihood estimate $\hat{\theta}^* = (V^{\mathrm{T}}W^{-1}V)^{-1}V^{\mathrm{T}}W^{-1}y$ with variance $V^{\mathrm{T}}W^{-1}V$ equal to the inverse of the second derivative. Also we see that $\ell^*(\theta)$ has the same linear and quadratic coefficients for θ as the full log-likelihood for the model $N(\theta V, W)$. Both $\ell_{ACL}(\theta)$ and $\ell^*(\theta)$ require variances and covariances of the score variables; see the Appendix.

Writing the uncorrected composite log-likelihood as $1^{\mathrm{T}}\underline{\ell}(\theta)$, where $\underline{\ell}(\theta)$ is the vector $\{\ell_1(\theta), \ldots, \ell_m(\theta)\}$, and $\ell_i(\theta) = -(1/2)v_{ii}\theta^2 + y_i\theta$, we can write

$$\mathrm{var}(\hat{\theta}_{UCL} - \hat{\theta}^*) = \frac{1^{\mathrm{T}}W1}{(1^{\mathrm{T}}V)^2} - \frac{1}{V^{\mathrm{T}}W^{-1}V},$$

expressing the first order loss of efficiency of the composite likelihood estimator relative to that obtained using the full first order log-likelihood.

5 Combining p-value functions

The p-value with its Uniform(0,1) distribution under the null provides a very clear presentation $p_i = \Phi\{(s_i - v_{ii}\theta)/v_{ii}^{1/2}\}$ of statistical inference from the i-th model with data. But an alternative presentation on the z scale using the related value $z_i = \Phi(p_i)$ gives an equivalent presentation with its Normal$(0, 1)$ distribution under the null; in fact this offers a clearer picture of how extreme an extreme value really is. A third presentation using the slope of the component log-likelihood $\ell_i'(\theta) = s_i - v_{ii}\theta$ needs an additional root information $v_{ii}^{1/2}$. The simple connections are

$$\begin{array}{lll} p_i & \longrightarrow z_i = \Phi^{-1}(p_i) & \longrightarrow (s_i - v_{ii}\theta) = (v_{ii})^{1/2} z_i, \\ p_i = \Phi(z_i) & \longleftarrow z_i = v_{ii}^{-1/2}(s_i - v_{ii}\theta) & \longleftarrow (s_i - v_{ii}\theta); \end{array}$$

thus we can move back and forth among the three types of presentation. But only one form of presentation has the additivity property that we have been investigating, and that form using score variables $s_i - v_{ii}\theta$ has available the $V^T W^{-1}$ technology for linearly combining log-likelihood information,

Now suppose we have available d component log-likelihoods and we wish to combine the related p-values. The resulting combined score is available from the expression following (8):

$$V^{\mathrm{T}} W^{-1}(s - V\theta) = V^{\mathrm{T}} W^{-1} V^{1/2} \Phi^{-1}\{p(\theta; s)\}, \tag{15}$$

where $V^{1/2}\Phi^{-1}\{p(\theta; s)\}$ is the vector with coordinates $v_{ii}^{1/2}\Phi^{-1}\{p(\theta; s_i)\}$. And then we can convert back obtaining the full first-order p-value for the combined data:

$$\tilde{p}(\theta; s) = \Phi[(V^T W^{-1} V)^{-1/2} V^{\mathrm{T}} W^{-1} V^{1/2} \Phi^{-1}\{p(\theta; s)\}]. \tag{16}$$

Example 2 continued: Dependent and exchangeable components. The composite score variable relative to the nominal parameter value $\theta_0 = 0$ is

$$V^{\mathrm{T}} W^{-1} s = \frac{2}{3}(y_1 + y_2),$$

which is the score variable from the new log-likelihood $\ell^*(\theta)$, and the relevant quantile is

$$z = \left(\frac{8}{3}\right)^{-1/2} \left\{\frac{2}{3}(y_1 + y_2) - \frac{8}{3}\theta\right\},$$

which has a standard normal distribution, exact in this case. The corresponding composite p-value function is then

$$\tilde{p}(\theta; s) = \Phi(z).$$

This p-value calculation extracts all the available information concerning θ.

Example 3 continued: Dependent but not exchangeable components. The combined score variable relative to the nominal $\theta_0 = 0$ is $V^{\mathrm{T}} W^{-1} s = y_2$, which is the score variable from the new log-likelihood $\ell^*(\theta)$. The corresponding quantile is $z = 2^{-1/2}(y_2 - 2\theta)$ and the corresponding composite p-value function is

$$\tilde{p}(\theta; s) = \Phi(z) = \Phi\{2^{-1/2}(y_2 - 2\theta)\},$$

which is in accord with the general observation that y_2 here provides full information on the parameter θ. This new p-value is fully accurate using all available information.

Example 7: Combining three p-values. Suppose three investigations of a common scalar parameter θ have yielded the following p-values for assessing a null value θ_0 : $1.15\%, 3.01\%.2.31\%$. For combining these we need the measures of precision as provided by the informations, say, $v_{11} = 3.0, v_{22} = 6.0, v_{33} = 9.0$. The corresponding z-values and score values s are

$$\begin{array}{ll} z_1 = \Phi^{-1}(0.0115) = -2.273 & s_1 - 3\theta_0 = 3^{1/2}(-2.273) = -3.938 \\ z_2 = \Phi^{-1}(0.0301) = -1.879 & s_2 - 6\theta_0 = 6^{1/2}(-1.879) = -4.603 \\ z_3 = \Phi^{-1}(0.0231) = -1.994 & s_3 - 9\theta_0 = 9^{1/2}(-1.994) = -5.981. \end{array} \qquad (17)$$

First suppose for simplicity that the investigations are independent, so that $W = \text{diag}(V)$ and $V^T W^{-1} = (1, 1, 1)$, which says that we should add the scores, as one would expect; this gives the combined score -14.522. Then standardizing by the root of the combined information $18^{1/2} = 4.243$ we obtain $\tilde{p} = \Phi(-3.423) = 0.00032$.

Now to examine the effect of dependence among the scores we consider a cross-correlation matrix of the form

$$R = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix},$$

with corresponding covariance matrix W having entries $3, 6, 9$ on the diagonal and the appropriate covariances otherwise. And then for a low-level of correlation we take $\rho = 0.2$. The coefficients for combining the scores s_i in the array (17) are given in the array

$$V^T W^{-1} = (3,6,9) \begin{pmatrix} 3.000 & 0.424 & 0.520 \\ 0.424 & 6.000 & 0.735 \\ 0.520 & 0.735 & 9.000 \end{pmatrix}^{-1}$$
$$= (3,6,9) \begin{pmatrix} 0.357 & -0.042 & -0.034 \\ -0.042 & 0.179 & -0.024 \\ -0.024 & 0.119 & 0.119 \end{pmatrix}$$

producing the row vector $(0.318, 0.643, 0.787)$ for combining the scores. The resulting z and p-value are -2.81 and $\tilde{p} = 0.0025$, a considerable increase from the independence value 0.00032.

6 Conclusion

In this paper we use likelihood asymptotics to construct a fully first-order accurate log-likelihood function for the composite likelihood context. It requires the covariance matrix of score variables, which is also needed for inference from composite likelihood.

The advantage of the first-order model is that it emphasizes that the model is equivalent under rotation to a sample from a location Normal with known variance. As such there are well substantiated procedures for analyzing a location θ with known variance and available likelihood.

The use of the first order model also leads to formulas for combining p-values, by converting them to score variables, then combining them using the vector $V^{\mathrm{T}} W^{-1}$, then converting back to the p-value scale. This gives widely applicable procedures for meta-analysis.

The methods of higher-order likelihood have been widely used for developing theory with third order accuracy for the analysis of statistical models with data. We have applied these methods here to the first order using the composite likelihood context, and have achieved corresponding resolutions with full first order accuracy. More generally, this first-order approach is available widely and can be applied more generally, in particular to contexts where estimating equations are used. This converts the related problems into the direct analysis of a first order model, thus simplifying and avoiding the use of estimating equations as an exploratory procedure.

Acknowledgements

Some portions of this material were presented at the University of Padova and University College London in 2012 and 2013. During the spring semester 2015 at the University of Toronto the material was developed and revised in a likelihood seminar and appreciation and thanks go to participants Wei Lin, Faizan Khalid Mohsin, Xin-Le Sheng, Harris Quach and Feng Xu.

This research has received support from the National Science and Engineering Research Council of Canada and the Senior Scholars Funding of York University.

References

Freedman, D. A. (2006). On the so-called "Huber sandwich estimator" and "Robust standard errors". *The American Statistician 60*, 299–302.

Jiang, W. and B. Turnbull (2004). The indirect method: inference based on intermediate statistics – a synthesis and examples. *Statistical Science 19*, 239 – 263.

Lindsay, B. (1988). Composite likelihood methods. In N. Prabhu (Ed.), *Statistical Inference from Stochastic Processes*, Volume 80, pp. 221 – 239. Providence, Rhode Island: American Mathematical Society.

Pace, L., A. Salvan, and N. Sartori (2011). Adjusting composite likelihood ratio statistics. *Statistica Sinica 21*, 129 – 148.

Pace, L., A. Salvan, and N. Sartori (2016). Combining dependent log likelihoods for a scalar parameter of interest. preprint.

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