Received: 10 September 2018 | Accepted: 14 November 2018
DOI: 10.1002/sta4.212

WILEY

ORIGINAL ARTICLE

Integrated likelihoods for functions of a parameter

Thomas A. Severini

Department of Statistics, Northwestern University, Illinois 60208, Evanston

Correspondence
Thomas A. Severini, Department of Statistics, Northwestern University, Evanston, Illinois 60208.
Email: severini@northwestern.edu

Non-Bayesian inference regarding a parameter of interest in the presence of a nuisance parameter may be based on the integrated likelihood function, in which the nuisance parameter is eliminated by averaging the likelihood function with respect to a weight function for the nuisance parameter. Recent research has shown that integrated likelihood methods work particularly well when the model is reparameterized in terms of a nuisance parameter chosen to be "unrelated" to the parameter of interest and the corresponding weight function for this nuisance parameter is chosen so that it does not depend on the parameter of interest. One choice for such a nuisance parameter is the *zero score expectation (ZSE)* parameter. The purpose of this note is to extend the definition of the ZSE parameter to the case in which the model has a parameter vector θ, with the parameter of interest of the model taken to be a function of θ; that is, the definition of the ZSE parameter is extended to the case in which there is not an explicit nuisance parameter for the model. The resulting integrated likelihood function has the same desirable properties as integrated likelihoods based on the ZSE parameter in models with an explicit nuisance parameter.

KEYWORDS

confidence intervals, integrated likelihood ratio statistic, likelihood inference, maximum integrated likelihood estimator, modified profile likelihood, orthogonal parameters, zero score expectation parameter

1 | INTRODUCTION

For inference in models with nuisance parameters, integrated likelihood functions provide a useful alternative to other likelihood-based methods, such as those based on the profile likelihood or marginal and conditional likelihood functions.

Consider a model with a scalar parameter of interest ψ, a d-dimensional nuisance parameter λ taking values in a set Λ, and a likelihood function $L(\psi, \lambda)$. An integrated likelihood function for ψ is of the form

$$\int_{\Lambda} L(\psi, \lambda)\pi(\lambda|\psi)d\lambda, \tag{1}$$

where π is a weight function on Λ.

Integrated likelihood methods are based on averaging the likelihood function over the set of possible values of the nuisance parameter and, hence, are closely related to the methods of Bayesian inference; see, for example, Berger, Liseo, and Wolpert (1999), Liseo (1993), and Liseo (2005) for discussions of integrated likelihoods from the Bayesian perspective. However, integrated likelihood methods also have a number of desirable properties from the non-Bayesian perspective; see, for example, Kalbfleisch and Sprott (1970) and Severini (2007, 2010, 2011). Furthermore, there is a growing list of examples in which integrated likelihood methods outperform more traditional likelihood-based methods; see, for example, Arellano and Bonhomme (2009), Berger et al. (1999), Carroll and Lombard (1985),Chamberlain (2007), De Bin, Sartori, and Severini (2015), Ghosh, Datta, Kim, and Sweeting (2006), Malley, Redner, Severini, Badner, Pajevic, and Bailey-Wilson (2003), Osborne and Severini (2000), and Schumann, Severini, and Tripathi (2018). This is particularly true in models in which the dimension of the nuisance parameter is large relative to the sample size (De Bin et al., 2015; Schumann et al., 2018).

Calculation of the integrated likelihood requires calculation, or approximation, of the integral appearing in (1). The nature of the terms appearing in the integral makes it well suited for some form of Monte Carlo integration. See Zhao and Severini (2017) for a discussion of different computational approaches and comparisons of various methods for computing an integrated likelihood.

Severini (2007) gives a detailed study of the selection of $\pi(\lambda|\psi)$ so that the resulting integrated likelihood is useful for non-Bayesian likelihood inference. This research identified a number of properties of an integrated likelihood as being important for non-Bayesian likelihood inference.

To simplify the exposition, for the remainder of this section, assume that the nuisance parameter is a scalar. Severini (2007) shows that the resulting integrated likelihood function has several desirable properties provided that $\pi(\lambda|\psi)$ is chosen so that, if γ is a nuisance parameter that is "unrelated" to ψ, then the corresponding implied weight function for γ under $\pi(\lambda|\psi)$, using the usual change-of-variable formula for density functions, should not depend on ψ. Here, we take "unrelated" to mean that $\hat{\gamma}_\psi$, the maximum likelihood estimator of γ for fixed ψ, is approximately constant as a function of ψ; specifically, we require that $\hat{\gamma}_\psi = \hat{\gamma} + O(|\psi - \hat{\psi}|n^{-\frac{1}{2}})$, where n denotes the sample size, in which case we say that γ is *strongly unrelated* to ψ.

For instance, using this approach, the resulting integrated likelihood function for ψ is approximately score- and information-unbiased (Severini, 2007); note that the desirable properties of the integrated likelihood function do not require a specific form for the weight function, provided that the implied weight function for γ does not depend on ψ.

Thus, one approach to choosing the weight function for an integrated likelihood function is to define a nuisance parameter ϕ that is strongly unrelated to ψ and then use a weight function for ϕ that does not depend on ψ. Note that an information-orthogonal nuisance parameter (Cox & Reid, 1987) is not, in general, strongly unrelated to ψ; thus, we consider the alternative approach to constructing a nuisance parameter ϕ, as proposed in Severini (2007).

Consider the implicit equation for ϕ given by

$$\mathrm{E}\{\ell_\lambda(\psi,\lambda);\hat{\psi},\phi\} \equiv \mathrm{E}\{\ell_\lambda(\psi,\lambda);\psi_0,\lambda_0\}|_{(\psi_0,\lambda_0)=(\hat{\psi},\phi)} = 0, \tag{2}$$

that is, fixing the value of $(\psi,\lambda,\hat{\psi})$ and solving (2) for ϕ yields $\phi(\psi,\lambda;\hat{\psi})$. Alternatively, given ϕ, the corresponding value of λ may be found by solving (2) for λ.

It may be shown that $\hat{\phi} = \hat{\lambda}$ and that $\hat{\phi}_\psi = \hat{\lambda} + O_p(n^{-\frac{1}{2}}|\hat{\psi} - \psi|)$ so that ϕ is strongly unrelated to ψ. The nuisance parameter ϕ is called the zero score expectation (ZSE) parameter. An important property of the ZSE parameter is that it depends on the data, through $\hat{\psi}$; however, such data dependence does not affect the properties of the integrated likelihood (Severini, 2007).

The likelihood function for (ψ,ϕ) may be written $L(\psi,\lambda(\psi,\phi))$, and hence, an integrated likelihood for ψ may now be formed by integrating $L(\psi,\lambda(\psi,\phi))$ with respect to $\pi(\phi)$:

$$\bar{L}(\psi) = \int_\Phi L(\psi,\lambda(\psi,\phi))\ \pi(\phi)d\phi,$$

where Φ is the space of possible ϕ. Note that for obtaining the integrated likelihood evaluated at ψ, it is important to be able to find the value of λ corresponding to given values of $\boldsymbol{\phi}$ and ψ (and $\hat{\psi}$) rather than to find the value of ϕ corresponding to a given value of λ.

The integrated likelihood, $\bar{L}(\psi)$, is shown to have many of the properties of a genuine likelihood function for ψ and, hence, is useful for non-Bayesian likelihood inference. Furthermore, $\bar{L}(\psi)$ is closely related to the modified profile likelihood, $L_M(\psi)$ with $\log\bar{L}(\psi) = \log L_M(\psi) + O(n^{-\frac{1}{2}})$ for fixed ψ and $\log\bar{L}(\psi) - \log\bar{L}(\hat{\psi}) = \log L_M(\psi) - \log L_M(\hat{\psi}) + O(n^{-1})$ for $\psi = \hat{\psi} + O(n^{-\frac{1}{2}})$; this result also establishes a connection between the integrated likelihood and marginal and conditional likelihood functions. See Severini (2007, 2010, 2011) for the properties of integrated likelihoods based on the ZSE parameterization for parameters of fixed dimension; De Bin et al. (2015) and Schumann et al. (2018) consider the case of *two-index asymptotics*, in which the dimension of the nuisance parameter increases with the sample size; two-index asymptotic captures the idea that the dimension of the nuisance parameter is large relative to the sample size.

The goal of this paper is to consider the construction of an integrated likelihood function for the case in which the model has parameter vector $\boldsymbol{\theta} = (\theta_1,\theta_2,\ldots,\theta_m)$, with the real-valued parameter of interest taken to be a function of $\boldsymbol{\theta}$: $\psi = g(\boldsymbol{\theta})$. That is, although the parameter of the model is the vector $\boldsymbol{\theta}$, the real-valued function $g(\boldsymbol{\theta})$ is of primary interest; the choice of the function $g(\cdot)$ for a particular analysis will depend on the context.

To construct an integrated likelihood in this setting, the definition of the ZSE parameter is extended to the case in which an explicit nuisance parameter is not available; that is, although the parameter of interest $g(\boldsymbol{\theta})$ is given, it is not necessary to find a function of $\boldsymbol{\theta}$ to serve as the nuisance parameter of the model. It is shown that the properties of the resulting integrated likelihood are identical to those of an ZSE-parameter-based integrated likelihood when there is an explicit nuisance parameter. This more general description of the ZSE parameterization extends the definition of the corresponding integrated likelihood function to a wider class of models.

2 | ZSE PARAMETERIZATION FOR FUNCTIONS OF θ

We now consider the extension of the ZSE parameter to the case in which the parameter of the model is a vector θ, taking values in a set Θ, with the parameter of interest $\psi = g(\theta)$. As is the case in the model parameterized by (ψ,λ), discussed in Section 1, in order to calculate the integrated likelihood, we need to be able to find the value of θ corresponding to a given value of the ZSE parameter.

In order to motivate the form of the ZSE parameter for the case in which an explicit nuisance parameter is not available, first suppose that the model is parameterized by (ψ,λ) for a given nuisance parameter λ so that an explicit nuisance parameter is available. Note that, although in Severini (2007) the value of λ corresponding to a given value of the ZSE parameter $\boldsymbol{\phi}$ was originally given as the solution in λ to Equation (2), it may be defined equivalently as

$$\arg\max_\lambda \mathrm{E}\{\ell(\psi,\lambda);\hat{\psi},\boldsymbol{\phi}\}; \tag{3}$$

see Section 4 of Schumann et al. (2018). The fact that λ satisfies (2) then follows under standard regularity conditions. Thus, we may express the relationship between the parameter θ and the ZSE parameter ϕ in terms of a constrained maximization problem.

Define the set

$$\Omega_{\hat{\psi}} = \{\omega \in \Theta : g(\omega) = \hat{\psi}\}. \tag{4}$$

It follows that elements of $\Omega_{\hat{\psi}}$ are of the form $(\hat{\psi}, \phi)$ where ϕ takes values in Λ, the set of possible λ.

Then, for a given value of $\omega \in \Omega_{\hat{\psi}}$, the corresponding value of θ may be described as the maximizer of

$$\mathrm{E}\,(\ell(\theta); \omega)$$

subject to the restriction that

$$g(\theta) = \psi.$$

That is, the value of $\theta = (\psi, \lambda)$ corresponding to $\omega \in \Omega_{\hat{\psi}}$ is obtained as the maximizer of

$$\mathrm{E}\,(\ell(\psi, \lambda); \omega)$$

with respect to λ. Here, $\mathrm{E}(\cdot; \theta)$ denotes the expectation operator corresponding to the distribution with parameter θ, and hence, $\mathrm{E}(\cdot; \omega)$ denotes the expectation operator corresponding to the distribution with the parameter θ taken to be ω.

This same approach can be used in the general case, in which θ is a vector parameter and the parameter of interest is given by $\psi = g(\theta)$, without an explicit nuisance parameter. Defining $\Omega_{\hat{\psi}}$ as in (4), for a given value of $\omega \in \Omega_{\hat{\psi}}$, the corresponding value of θ may be described as the maximizer of

$$\mathrm{E}\,(\ell(\theta); \omega)$$

subject to the restriction that

$$g(\theta) = \psi.$$

This characterization yields a function from $\Omega_{\hat{\psi}}$ to the set

$$\{\theta \in \Theta : g(\theta) = \psi\}; \tag{5}$$

note that elements of the set (5) are of the form (ψ, λ), where λ is an appropriate nuisance parameter. However, explicit construction of such a nuisance parameter is not needed. Thus, ω plays the role of the ZSE parameter, which, in this context, is subject to the restriction that $g(\omega) = \hat{\psi}$.

For instance, if the constrained maximization problem is solved using Lagrange multipliers, then the relationship is given by solving the equations

$$\begin{aligned} \mathrm{E}\left(\ell_{\theta_j}(\theta); \omega\right) + \alpha g_{\theta_j}(\theta) = 0, \quad j = 1, 2, \ldots, m \\ g(\theta) = \psi. \end{aligned}$$

Thus, we may write $\theta = b(\omega, \psi; \hat{\psi})$ for some function b; it follows that the likelihood function for ϕ is given by $L(b(\omega, \psi; \hat{\psi})$. Note that, for each $\omega \in \Omega_{\hat{\psi}}$,

$$g(b(\omega, \psi; \hat{\psi})) = \psi.$$

The corresponding integrated likelihood function evaluated at ψ is obtained by integrating $L(b(\omega, \psi; \hat{\psi}))$ with respect to ω over the set $\Omega_{\hat{\psi}}$.

An important property of the ZSE parameter, as defined in Severini (2007), is that it is invariant under interest-respecting reparameterizations; see Severini (2007) and Schumann et al. (2018). This same property applies in the present context. This result implies that the integrated likelihood function based on the ZSE parameter ω as defined in this section has the same properties as the integrated likelihood function based on the parameterization (ψ, λ) for an appropriate nuisance parameter λ. It follows that the integrated likelihood function based on ϕ has the same properties as the ZSE-based likelihood discussed in Severini (2007, 2010), De Bin et al. (2015), and Schumann et al. (2018). Technical details regarding these results are given in the Appendix.

3 | AN EXAMPLE

We now illustrate this approach on the following example. Consider independent random variables Y_{jk}, $k = 1, 2, \ldots, n_j$, $j = 1, 2, \ldots, m$, such that Y_{jk} has an exponential distribution with rate parameter θ_j so that the marginal distribution of Y_{jk} has density $\theta_j \exp(-\theta_j y)$, $y > 0$. It follows that the log-likelihood function for parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ is given by

$$\ell(\theta) = -\sum_{j=1}^{m} n_j \log(\theta_j) - \sum_{j=1}^{q} \theta_j Y_j, \quad \theta_j > 0, \quad j = 1, 2, \ldots, m,$$

where

$$Y_j = \sum_{k=1}^{n_j} Y_{jk}.$$

Suppose that we are interested in the variability of the values in the sequence $\theta_1, \theta_2, \dots, \theta_m$. Thus, consider the parameter of interest $\psi = g(\theta)$ given by

$$g(\boldsymbol{\theta}) = \left(\frac{1}{m} \sum_{j=1}^{m} (\theta_j - \bar{\theta})^2 \right)^{\frac{1}{2}}, \quad \bar{\theta} = \frac{1}{m} \sum_{j=1}^{m} \theta_j.$$

To calculate the integrated likelihood for ψ, we need to find the value of $\boldsymbol{\theta}$ corresponding to specific values of $\boldsymbol{\omega}, \psi$, and $\hat{\psi}$, as described in Section 2. Here,

$$\mathrm{E}\,(\ell(\boldsymbol{\theta}); \boldsymbol{\omega}) = -\sum_{j=1}^{m} n_j \log(\theta_j) - \sum_{j=1}^{m} n_j \theta_j / \omega_j;$$

hence, for $\boldsymbol{\omega}$ satisfying

$$\frac{1}{m} \sum_{j=1}^{m} (\omega_j - \bar{\omega})^2 = \hat{\psi}^2,$$

we find $\theta_1, \theta_2, \dots, \theta_m$ that solve the constrained maximization problem

$$\max -\sum_{j=1}^{m} n_j \log(\theta_j) - \sum_{j=1}^{m} n_j \theta_j / \omega_j \quad \text{subject to} \quad \frac{1}{m} \sum_{j=1}^{m} (\theta_j - \bar{\theta})^2 = \psi^2. \tag{6}$$

Although this problem is difficult to solve analytically, it is easily solved numerically, given values for $\omega_1, \omega_2, \dots, \omega_m$, $\hat{\psi}$, and ψ.

To calculate the integrated likelihood function for ψ, any one of a number of Monte Carlo methods can be used (Zhao & Severini, 2017). For simplicity, here, we describe what Zhao and Severini (2017) refer to as "simple Monte Carlo"; this approach also has the benefit of highlighting the relationship between the integrated likelihood $\bar{L}(\psi)$ and the profile likelihood $L_p(\psi)$. Of course, more sophisticated (and efficient) Monte Carlo methods may also be used.

Values of $\omega_1, \omega_2, \dots, \omega_m$ are drawn from a given distribution, based on the weight function used to define the integrated likelihood; as noted previously, the properties of the integrated likelihood do not depend on the weight function used, provided that it does not depend on ψ. Furthermore, the weight function can depend on the data.

Note that the integral is over the set $\Omega_{\hat{\psi}}$, and hence, the values $\tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_m$ chosen must satisfy the condition that

$$\frac{1}{m} \sum_{j=1}^{m} \tilde{\omega}_j^2 - \left(\frac{1}{m} \sum_{j=1}^{m} \tilde{\omega}_j \right)^2 = \hat{\psi}^2. \tag{7}$$

In the present example, one approach is to use a distribution under which the ω_j are independent and then rescale the values so that (7) holds.

For each set of random variates $\tilde{\omega}_1^{(i)}, \tilde{\omega}_2^{(i)}, \dots, \tilde{\omega}_m^{(i)}$, the constrained maximization problem given by (6) is solved for $\tilde{\theta}_1^{(1)}, \tilde{\theta}_2^{(i)}, \dots, \tilde{\theta}_m^{(i)}$ and an estimate of $\bar{L}(\psi)$ is given by

$$\frac{1}{I} \sum_{i=1}^{I} L(\tilde{\boldsymbol{\theta}}^{(i)}),$$

where $\tilde{\boldsymbol{\theta}}^{(i)} = (\tilde{\theta}_1^{(1)}, \tilde{\theta}_2^{(i)}, \dots, \tilde{\theta}_m^{(i)})$ and I is the number of Monte Carlo replications used.

It is interesting to note that calculation of the profile likelihood has a very similar form. To calculate $L_p(\psi)$, we again solve the constrained maximization problem (6), but with ω_j taken to be the maximum likelihood estimator of θ_j, $\hat{\theta}_j = n_j / Y_j$; note that $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m) \in \Phi_{\hat{\psi}}$. The solution to (6) yields $\hat{\boldsymbol{\theta}}_\psi$, the constrained maximum likelihood estimator of $\boldsymbol{\theta}$ and, hence,

$$L_p(\psi) = L(\hat{\boldsymbol{\theta}}_\psi).$$

Thus, the integrated likelihood function based on the present approach may be viewed as modification of the profile likelihood based on incorporating additional variability into the calculation and then averaging the results. This additional variability addresses the fact that the profile likelihood is "overly optimistic," in the sense that it uses the value of the nuisance parameter that yields the greatest value of the likelihood function, for a given value of the parameter of interest.

Consider the data in Table 1 of Proschan (1963), on the intervals between failures of the air conditioning equipment on Boeing 720 aircraft. The analysis in Proschan (1963) shows that the times between failures for the different aircraft appear to be exponentially distributed, but the failure rates vary among the aircraft. Proschan finds that this variability in the failure rates leads to an observed decreasing failure rate when the data from the different aircraft are combined into a single set of data. Thus, it is of interest to measure the variability in the aircrafts' failure rates.

Data on 13 aircraft are given; however, four of the aircraft underwent "major overhauls" so that, following Proschan (1963), we analyse these as separate aircraft. Thus, there are data on 17 aircraft, with sample sizes (in order),

2, 3, 3, 5, 6, 6, 9, 10, 11, 12, 12, 13, 16, 24, 24, 27, 30.

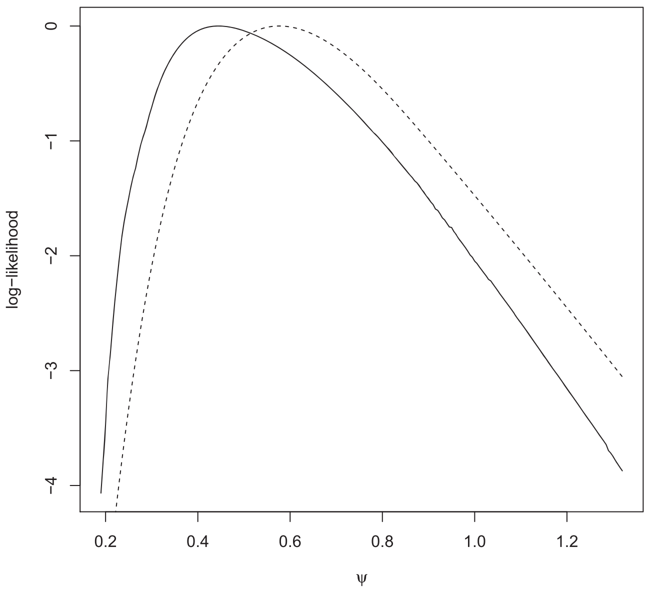

FIGURE 1 Integrated Likelihood Function in the Example

Figure 1 contains a plot of the integrated log-likelihood function, using the methodology described in this section, together with the profile log-likelihood (as a dashed line); both log-likelihood functions have been standardized to have maximum value 0. That is, each log-likelihood is standardized by subtracting the maximum value of the log-likelihood.

Note that the profile log-likelihood has the same basic shape as the integrated log-likelihood, but it is shifted to the right. The maximum integrated log-likelihood estimate of θ is 0.441, whereas the maximum likelihood estimate is 0.577. The 95% integrated likelihood ratio confidence interval, based on appproximating the distribution of the integrated likelihood ratio statistic by a chi-squared distribution with one degree of freedom, is given by $[0.233, 0.978]$; the analogous likelihood ratio confidence interval, based on the profile likelihood function, is given by $[0.309, 1.092]$.

To compare the properties of the integrated likelihood ratio and likelihood ratio confidence intervals, a small simulation study was conducted. Data were simulated from the exponential distributions using the same sample sizes as in the aircraft air conditioning data, with the parameter values taken to be the maximum likelihood estimates based on these data; thus, the true value of ψ is then 0.577.

Using a Monte Carlo sample size of 2,000, the coverage probabilities of the 95% integrated likelihood ratio and likelihood ratio confidence intervals were estimated; based on these results, the integrated likelihood ratio confidence interval has a coverage probability of 0.939, whereas the likelihood ratio confidence interval has a coverage probability of 0.847. These estimates have Monte Carlo standard errors of approximately 0.0054 and 0.0080, respectively. It follows that, although the coverage probability of the integrated likelihood ratio confidence interval is slightly less than its nominal value, it is a great improvement over the coverage probability of the standard likelihood ratio confidence interval; this agrees with the theoretical results obtained in Severini (2010).

ORCID

Thomas A. Severini https://orcid.org/0000-0001-6216-9416

REFERENCES

Arellano, M., & Bonhomme, S. (2009). Robust priors in nonlinear panel data models. *Econometrica*, *77*, 489–536.

Berger, J. O., Liseo, B., & Wolpert, R. (1999). Integrated likelihood functions for eliminating nuisance parameters (with discussion). *Statistical Science*, *14*, 1–28.

Carroll, R. J., & Lombard, F. (1985). A note on *n* estimators for the binomial distribution. *Journal of the American Statistical Association*, *80*, 423–426.

Chamberlain, G. (2007). Decision theory applied to an instrumental variables model. *Econometrica*, *75*, 609–52.

Cox, D. R., & Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *49*, 1–39.

De Bin, R., Sartori, N., & Severini, T. A. (2015). Integrated likelihoods in models with stratum nuisance parameters. *Electronic Journal of Statistics*, *9*, 1474–1491.

Ghosh, M., Datta, G. S., Kim, D., & Sweeting, T. J. (2006). Likelihood-based inference for ratios of regression coefficients in linear models. *Annals of the Institute of Statistical Mathematics*, *58*, 457–73.

Kalbfleisch, J. D., & Sprott, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *32*, 175–208.

Liseo, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika*, *80*, 295–304.

Liseo, B. (2005). The elimination of nuisance parameters, *Bayesian Thinking: Modeling and Computation* (pp. 193–219), Handbook of Statistics, Vol. *25*. Amersterdam: Elsevier/North Holland.

Malley, J. D., Redner, R. A., Severini, T. A., Badner, J. A., Pajevic, S., & Bailey-Wilson, J. E. (2003). Estimation of linkage and association from allele transformation data. *Biometrical Journal*, *45*, 349–66.

Osborne, J. A., & Severini, T. A. (2000). Inference for exponential order statistics models based on an integrated likelihood function. *Journal of the American Statistical Association*, *95*, 1220–1228.

Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics*, *5*, 375–383.

Schumann, M., Severini, T. A., & Tripathi, G. (2018). Integrated likelihood based inference for nonlinear panel data models with unobserved effects. Submitted.

Severini, T. A. (2007). Integrated likelihood functions for non-Bayesian inference. *Biometrika*, *94*, 529–42.

Severini, T. A. (2010). Likelihood ratio statistics based on an integrated likelihood. *Biometrika*, *97*, 481–496. https://doi.org/10.1093/biomet/asq015

Severini, T. A. (2011). Frequency properties of inferences based on an integrated likelihood function. *Statistica Sinica*, *21*(1), 433–447.

Zhao, Z., & Severini, T. A. (2017). Integrated likelihood computation methods. *Computational Statistics*, *32*, 281–313.

How to cite this article: Severini TA. Integrated likelihoods for functions of a parameter. *Stat*. 2018;7:e212. https://doi.org/10.1002/sta4.212

APPENDIX A

First consider the invariance of the ZSE parameter $\boldsymbol{\omega}$, as defined in Section 2, under interest-respecting reparameterizations of the model.

Let $\boldsymbol{\eta}$, taking values in a set $\mathcal{H}$, denote an alternative parameter of the original model; thus, $\boldsymbol{\eta} = q(\theta)$ and $\theta = Q(\boldsymbol{\eta})$ for some one-to-one functions $q(\cdot)$ and $Q(\cdot)$. Let h denote a function on the range of $\boldsymbol{\eta}$ satisfying $h(\boldsymbol{\eta}) = g(Q(\boldsymbol{\eta}))$ so that the parameter of interest for the model may be written $\psi = h(\boldsymbol{\eta})$.

For given values of ψ and $\hat{\psi}$, let $\boldsymbol{\omega}^*$ be an element of $\Omega_{\hat{\psi}}$ and let θ^* denote the corresponding value of the parameter θ. That is, θ^* solves the constrained maximization problem

$$\max \ \mathrm{E}\left(\ell(\theta); \boldsymbol{\omega}^*\right)$$

subject to

$$g(\theta) = \psi.$$

Let $\boldsymbol{\gamma}^* = q(\boldsymbol{\omega}^*)$; then

$$\boldsymbol{\gamma}^* \in \Gamma_{\hat{\psi}},$$

where

$$\Gamma_{\hat{\psi}} = \{\boldsymbol{\gamma} \in \Gamma : h(\boldsymbol{\gamma}) = \hat{\psi}\}.$$

Let $\boldsymbol{\eta}^*$ denote the value of the parameter $\boldsymbol{\eta}$ corresponding to $\boldsymbol{\gamma}^*, \psi$, and $\hat{\psi}$. That is, $\boldsymbol{\eta}^*$ solves the constrained maximization problem

$$\max \ \tilde{\mathrm{E}}\left(\tilde{\ell}(\boldsymbol{\eta}); \boldsymbol{\gamma}_1\right) \quad \text{subject to} \quad h(\boldsymbol{\eta}) = \psi.$$

Here, $\tilde{\ell}(\cdot)$ denotes the log-likelihood in the η-parameterization and $\tilde{\mathrm{E}}$ denotes the expectation operator in the η-parameterization, so that

$$\tilde{\ell}(\boldsymbol{\eta}) = \ell(Q(\boldsymbol{\eta})) \quad \text{and} \quad \tilde{\mathrm{E}}(\cdot; \boldsymbol{\eta}) = \mathrm{E}(\cdot; Q(\boldsymbol{\eta})).$$

Note that, for any $\theta_1 \in \Theta$, $g(\theta_1) = \psi$ if and only if $h(\boldsymbol{\eta}_1) = \psi$, where $\boldsymbol{\eta}_1 = q(\theta_1)$. Furthermore, using the fact that $\theta_1 = Q(\boldsymbol{\eta}_1)$,

$$\mathrm{E}\left(\ell(\theta_1); \boldsymbol{\omega}^*\right) = \mathrm{E}\left(\ell(Q(\boldsymbol{\eta}_1); Q(\boldsymbol{\gamma}^*)\right) = \tilde{\mathrm{E}}\left(\tilde{\ell}(\boldsymbol{\eta}_1); \boldsymbol{\gamma}^*\right).$$

It follows that $\theta^* = Q(\boldsymbol{\eta}^*)$.

This result can now be used to describe the relationship between the ZSE-parameter-based integrated likelihood functions for ψ using the two parameterizations. The fact that $\theta^* = Q(\boldsymbol{\eta}^*)$ shows that

$$L(\theta^*) = L(Q(\boldsymbol{\eta}^*)) = \tilde{L}(\boldsymbol{\eta}^*),$$

where $\tilde{L}(\cdot)$ denotes the likelihood function for $\boldsymbol{\eta}$. The integrated likelihoods are formed by integrating L and $\tilde{L}$ over the sets $\Omega_{\hat{\psi}}$ and $\Gamma_{\hat{\psi}}$, respectively, with respect to a weight function for the ZSE parameter. Hence, the integrated likelihood functions for ψ based on the different parameterizations do not necessarily agree unless the weight function for $\boldsymbol{\gamma}$, $\tilde{\pi}(\cdot)$, satisfies

$$\tilde{\pi}(\boldsymbol{\gamma}) = \pi(Q(\boldsymbol{\gamma})) \ |J_Q(\boldsymbol{\gamma})|,$$

where $\pi(\cdot)$ denotes the weight function for $\boldsymbol{\omega}$ and J_Q denotes the Jacobian of the function $Q(\cdot)$.

However, provided that neither weight function depends on ψ, the asymptotic properties of the two integrated likelihoods, as discussed in Severini (2007, 2010), De Bin et al. (2015), and Schumann et al. (2018), will be the same.

Let (ψ, λ) denote a parameter for the model such that the first component of the parameter is the parameter of interest and let $\boldsymbol{\phi}$ denote the ZSE parameter, as defined in Severini (2010). The results in this appendix show that the integrated likelihood function based on the ZSE parameter $\boldsymbol{\omega}$ defined in Section 2 has the same asymptotic properties as the integrated likelihood function based on the ZSE parameter $\boldsymbol{\phi}$.