

Parameter transformations and nonparametric likelihood inference

Trasformazioni parametriche e inferenza basata sulla verosimiglianza non parametrica

Thomas J. DiCiccio
Department of Statistical Science
Cornell University
tjd9@cornell.edu

Riassunto: Questo lavoro riguarda l'inferenza non parametrica su un parametro scalare stimato mediante una funzione di uno stimatore M multivariato. Gli intervalli asintotici, generalmente, risultano piuttosto diversi dagli intervalli ottenuti mediante verosimiglianze non parametriche. Il lavoro propone una trasformazione del parametro, tale che l'intervallo di confidenza asintotico approssimi meglio quelli basati su funzioni di verosimiglianza non parametriche. Le proprietà di questa trasformazione sono analizzate sia sotto l'aspetto teorico sia attraverso esempi numerici.

Keywords: asymptotics, bootstrap, confidence interval, M -estimation, skewness.

1. A common inference problem

Let X be a random vector having distribution function F , and let $\theta = (\theta^1, \dots, \theta^k)'$ be $\theta(F)$ be a functional whose value is determined by the equation $E_F\{\psi(X, \theta)\} = 0$, where

$$\psi(X, \theta) = \{\psi_1(X, \theta), \dots, \psi_k(X, \theta)\}'.$$

Given a random sample $X_1, \dots, X_n$ drawn from F , let $\hat{\theta}$ denote the M -estimator of θ defined by the equation

$$\sum_{i=1}^n \psi(X_i, \hat{\theta}) = 0.$$

Under standard regularity conditions (Huber, 1981), the quantity $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically normally distributed with mean zero and variance matrix $V = A^{-1}\Sigma(A^{-1})'$, where $A = (A_{a/b})$ and $\Sigma = (\Sigma_{ab})$ are given by

$$A_{a/b} = E\{\psi_{a/b}(X, \theta)\}, \quad \Sigma_{ab} = E\{\psi_a(X, \theta)\psi_b(X, \theta)\},$$

with $\psi_{a/b}(X, \theta) = \partial\psi_a(X, \theta)/\partial\theta^b$ ($a, b = 1, \dots, k$).

Suppose that the parameter of interest $\gamma = g(\theta)$ is a scalar function of θ , and let γ be estimated by $\hat{\gamma} = g(\hat{\theta})$. The asymptotic distribution of $\sqrt{n}(\hat{\gamma} - \gamma)$ is $N(0, \sigma^2)$, where

$$\sigma^2 = g'_\theta V g_\theta,$$

and $g_\theta = \partial g / \partial \theta$.

Inference for γ is usually achieved by using a studentized statistic $s_\gamma = \sqrt{n}(\hat{\gamma} - \gamma)/\hat{\sigma}$, where $\hat{\sigma}^2$ is a consistent estimator of σ^2 . A widely used method for constructing $\hat{\sigma}^2$ is to estimate A and Σ by their sample versions $\hat{A} = (\hat{A}_{a/b})$ and $\hat{\Sigma} = (\hat{\Sigma}_{ab})$, where

$$\hat{A}_{a/b} = n^{-1} \sum_{i=1}^n \psi_{a/b}(X_i, \hat{\theta}), \quad \hat{\Sigma}_{ab} = n^{-1} \sum_{i=1}^n \psi_a(X_i, \hat{\theta}) \psi_b(X_i, \hat{\theta});$$

then, the covariance matrix of $\sqrt{n}(\hat{\theta} - \theta)$ is estimated by $\hat{V} = \hat{A}^{-1} \hat{\Sigma} (\hat{A}^{-1})'$ and $\hat{\sigma}^2 = \hat{g}'_\theta \hat{V} \hat{g}_\theta$, where $\hat{g}_\theta = g_\theta(\hat{\theta})$.

Uncertainty about γ is frequently summarized in a symmetric two-sided approximate confidence interval having nominal coverage $100(1 - 2\alpha)\%$. If the distribution of s_γ is approximated by the $N(0, 1)$ distribution, or if the distribution of s_γ^2 is approximated by the χ_1^2 distribution, then the resulting interval is

$$\hat{\gamma} \pm n^{-1/2} z_\alpha \hat{\sigma}, \quad (1)$$

where z_α is the $100(1 - \alpha)$ percentile of the $N(0, 1)$ distribution. Since the chi-squared approximation to the distribution of s_γ^2 has error of order $O(n^{-1})$, this approximate interval has coverage error of order $O(n^{-1})$.

There are other methods for constructing approximate confidence intervals having coverage error of order $O(n^{-1})$. One method that has received considerable attention recently is the empirical likelihood of Owen (1988, 1990).

2. Nonparametric likelihood

For a given value of θ , the empirical likelihood $L(\theta)$ is defined to be the maximum of $\prod_{i=1}^n p_i(\theta)$ over all probability distributions $\{p_1(\theta), \dots, p_n(\theta)\}$ satisfying the constraints

$$\sum_{i=1}^n \psi_a(X_i, \theta) p_i(\theta) = 0 \quad (a = 1, \dots, k).$$

It can be shown that the maximum is attained when

$$p_i(\theta) = \frac{1}{n\{1 + \nu(\theta)' \psi(X_i, \theta)\}} \quad (i = 1, \dots, n),$$

and the vector of Lagrange multipliers $\nu(\theta) = \{\nu^1(\theta), \dots, \nu^k(\theta)\}'$ is determined by the given constraints. Thus, the empirical log likelihood function is

$$l(\theta) = - \sum_{i=1}^n \log\{1 + \nu(\theta)' \psi(X_i, \theta)\},$$

where

$$\sum_{i=1}^n \frac{\psi_a(X_i, \theta)}{1 + \nu(\theta)' \psi(X_i, \theta)} = 0 \quad (a = 1, \dots, k), \quad \nu(\theta)' \psi(X_i, \theta) \geq (1-n)/n \quad (i = 1, \dots, n).$$

Since $\nu(\hat{\theta}) = 0$, it follows that $l(\theta)$ has the property $l(\hat{\theta}) = 0$.

For given value of the interest parameter γ , let $\hat{\theta}_\gamma$ be the point that maximizes the empirical log likelihood function $l(\theta)$ subject to the constraint $g(\theta) = \gamma$. In particular, $\hat{\theta}_{\hat{\gamma}} = \hat{\theta}$. The profile empirical log likelihood function $l_p(\gamma)$ is defined by $l_p(\gamma) = l(\hat{\theta}_\gamma)$, and the empirical likelihood ratio statistic for γ is

$$w_\gamma = 2\{l_p(\hat{\gamma}) - l_p(\gamma)\} = -2l_p(\gamma) = -2l(\hat{\theta}_\gamma).$$

The profile empirical log likelihood is quite demanding to compute, so the associated likelihood ratio statistic can be burdensome to implement in practice.

A key result is that, as in the parametric framework, w_γ is asymptotically distributed as χ_1^2 , and the error in this approximation is typically of order $O(n^{-1})$. Consequently, the set

$$\{\gamma | w_\gamma \leq z_\alpha^2\} \quad (2)$$

is an approximate confidence interval having coverage $1 - 2\alpha + O(n^{-1})$.

3. Asymptotic expansions

Let $\hat{\eta} = (\hat{A}^{-1})^{-1}\hat{g}_\theta$ and $\hat{\delta} = -\hat{V}\hat{g}_\theta$. The usual empirical influence function at X_i is $EIF(X_i) = -\hat{\eta}'\psi(X_i, \theta)$ ($i = 1, \dots, n$), and $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n \{EIF(X_i)\}^2$. Now, define the quantity $\hat{k}_3$ by

$$\begin{aligned} \hat{k}_3 = & n^{-1/2} \left[2n^{-1} \sum_{i=1}^n \{EIF(X_i)\}^3 + 6 \frac{d}{d\epsilon} n^{-1} \sum_{i=1}^n EIF(X_i) \hat{\eta}' \psi(X_i, \hat{\theta} + \epsilon \hat{\delta}) \right. \\ & \left. + 3 \frac{d^2}{d\epsilon^2} n^{-1} \sum_{i=1}^n \hat{\eta}' \psi(X_i, \hat{\theta} + \epsilon \hat{\delta}) - 3 \frac{d^2}{d\epsilon^2} g(\hat{\theta} + \epsilon \hat{\delta}) \right] \Big|_{\epsilon=0} / \hat{\sigma}^3, \end{aligned}$$

which is of order $O_p(n^{-1/2})$. The derivatives appearing in $\hat{k}_3$ can be evaluated by using the standard approximations to the first- and second-order derivatives of a real-valued function $f(\hat{\theta} + \epsilon \lambda)$ at $\epsilon = 0$:

$$\begin{aligned} \frac{d}{d\epsilon} f(\hat{\theta} + \epsilon \lambda) \Big|_{\epsilon=0} & \approx \frac{f(\hat{\theta} + \Delta \lambda) - f(\hat{\theta} - \Delta \lambda)}{2\Delta}; \\ \frac{d^2}{d\epsilon^2} f(\hat{\theta} + \epsilon \lambda) \Big|_{\epsilon=0} & \approx \frac{f(\hat{\theta} + \Delta \lambda) - 2f(\hat{\theta}) + f(\hat{\theta} - \Delta \lambda)}{\Delta^2}, \end{aligned}$$

where $\Delta = h(\lambda' \hat{V} \lambda)^{-1/2}$ and h is chosen to be small, say between 0.001 and 0.01.

DiCiccio & Monti (2001) showed that the first three derivatives of the profile empirical log likelihood function satisfy:

$$l_p^{(1)}(\hat{\gamma}) = 0, \quad l_p^{(2)}(\hat{\gamma}) = -n/\hat{\sigma}^2, \quad l_p^{(3)}(\hat{\gamma}) = -n^{3/2} \hat{k}_3 / \hat{\sigma}^3.$$

It follows then that the profile empirical likelihood ratio statistic w_γ has the following expansion in terms of the studentized quantity s_γ :

$$w_\gamma = s_\gamma^2 - \frac{1}{3} \hat{k}_3 s_\gamma^3 + O_p(n^{-1}).$$

Moreover, the endpoints of confidence interval (2) obtained by using the χ_1^2 approximation to the distribution of w_γ are

$$\hat{\gamma} \pm n^{-1/2} \hat{\sigma} z_\alpha - n^{-1/2} \hat{\sigma} \frac{1}{6} \hat{k}_3 z_\alpha^2 + O_p(n^{-3/2}). \quad (3)$$

Comparison of (1) and (3) shows that the endpoints of the standard intervals differ from the endpoints of the empirical likelihood intervals by the amount $n^{-1/2} \hat{\sigma} \frac{1}{6} \hat{k}_3 z_\alpha^2 + O_p(n^{-3/2})$, which is of order $O_p(n^{-1})$.

3. Parameter transformations

The preceding expansions show that the symmetric interval (1) is close to the empirical likelihood interval (2) when the symmetric interval is computed in terms of a parameterization for which $\hat{k}_3$ is small. In particular, if $\hat{k}_3 = 0$, then the difference between the endpoints of these intervals is of order $O_p(n^{-3/2})$. It is useful then to identify parameterizations for which $\hat{k}_3 = 0$.

Let $\tau = t(\gamma)$ be a monotonically increasing transformation of the interest parameter, and let $l_{p,\gamma}(\gamma)$ and $l_{p,\tau}(\tau)$ denote the profile empirical log likelihood functions for γ and τ , respectively. Differentiation of the identity $l_{p,\gamma}(\gamma) = l_{p,\tau}\{t(\gamma)\}$ shows that

$$l_{p,\tau}^{(3)}(\hat{\gamma}) \{t^{(1)}(\hat{\gamma})\}^3 = l_{p,\gamma}^{(3)}(\hat{\gamma}) - 3l_{p,\gamma}^{(2)}(\hat{\gamma}) \{t^{(2)}(\hat{\gamma})/t^{(1)}(\hat{\gamma})\},$$

and hence,

$$\hat{k}_{3,\tau} = \hat{k}_{3,\gamma} - 3n^{-1/2} \hat{\sigma}_\gamma \{t^{(2)}(\hat{\gamma})/t^{(1)}(\hat{\gamma})\}.$$

Thus, the condition $\hat{k}_{3,\tau} = 0$ is achieved when the transformation $\tau = t(\gamma)$ satisfies

$$d \log \{t^{(1)}(\gamma)\} / d\gamma \big|_{\gamma=\hat{\gamma}} = \frac{1}{3} n^{1/2} \hat{k}_{3,\gamma} / \hat{\sigma}_\gamma. \quad (4)$$

Suppose that the interest parameter γ is positive, and consider the Box-Cox class of transformations $t(\gamma) = (\gamma^\lambda - 1)/\lambda$. Then, $d \log \{t^{(1)}(\gamma)\} / d\gamma = (\lambda - 1)/\gamma$, and equation (4) yields

$$\lambda = 1 + \frac{1}{3} n^{1/2} \hat{\gamma} \hat{k}_{3,\gamma} / \hat{\sigma}_\gamma. \quad (5)$$

On the other hand, if there are no restrictions on the range of γ , then the class of transformations $t(\gamma) = \{\exp(\lambda\gamma) - 1\}/\lambda$ can be entertained. This class has $d \log \{t^{(1)}(\gamma)\} / d\gamma = \lambda$, and equation (4) yields

$$\lambda = \frac{1}{3} n^{1/2} \hat{k}_{3,\gamma} / \hat{\sigma}_\gamma. \quad (6)$$

It is crucial to keep in mind that the parameterizations determined by equations (5) and (6) are data-dependent, and as such, they have no direct interpretation. The data-dependent parameterizations are introduced here merely as a device for making the standard symmetric intervals (1) resemble more closely empirical likelihood intervals; to obtain interpretable results, it would be necessary in practice to transform the resulting intervals back to the original scale. Back transforming is easy, however, since the inverse of the Box-Cox transformation has a closed form expression. The back-transformed intervals would retain the desirable property of being nearly empirical likelihood intervals, but their computation would retain the simplicity of the standard intervals.

4. Interpretation of the parameter transformations

DiCiccio & Monti (2002) showed that the skewness of the studentized statistic $s_\gamma = \sqrt{n}(\hat{\gamma} - \gamma)/\hat{\sigma}$ is

$$k_3 = n^{-1/2} E \left(\left[2\{\eta' \psi(X, \theta)\}^3 + 6\{\eta' \psi(X, \theta)\} \frac{d}{d\epsilon} \eta' \psi(X, \theta + \epsilon \delta) + 3 \frac{d^2}{d\epsilon^2} \eta' \psi(X, \theta + \epsilon \delta) - 3 \frac{d^2}{d\epsilon^2} g(\theta + \epsilon \delta) \right] \Big|_{\epsilon=0} \right) / \sigma^3,$$

to error of order $O(n^{-1})$, where $\eta = (A^{-1})' g_\theta$ and $\delta = -V g_\theta$. Comparison of k_3 with the quantity $\hat{k}_3$ shows that $\hat{k}_3 = k_3 + O_p(n^{-1})$.

If $\tau = t(\gamma)$ is a monotonically increasing transformation, then the studentized statistic, expressed in terms of the parameter τ is

$$s_\tau = \sqrt{n}(\hat{\tau} - \tau)/\sigma_\tau = \sqrt{n}\{t(\hat{\gamma}) - t(\gamma)\}/\{\sigma_\gamma t^{(1)}(\hat{\gamma})\}.$$

Let $k_{3,\gamma}$ and $k_{3,\tau}$ denote the versions of k_3 for s_γ and s_τ , respectively. Then,

$$k_{3,\tau} = k_{3,\gamma} - 3n^{-1/2} \sigma_\gamma \{t^{(2)}(\gamma)/t^{(1)}(\gamma)\}.$$

It follows that a transformation $\tau = t(\gamma)$ reduces the skewness of the studentized statistic to order $O(n^{-1})$ provided

$$d \log \{t^{(1)}(\gamma)\} / d\gamma = \frac{1}{3} n^{1/2} k_{3,\gamma} / \sigma_\gamma.$$

Thus, the transformations introduced to symmetrize the profile empirical log likelihood function can be viewed as estimated versions of transformations that reduce the skewness of the studentized statistic.

The connections between symmetry of the profile log likelihood and reduced skewness of the studentized statistic have been investigated for parametric models by Sprott (1980). The results here extend those connections to the nonparametric framework. See also DiCiccio & Tibshirani (1987) and Konishi (1991). The transformations presented in Section 4 were developed from the point of view of skewness reduction by DiCiccio & Monti (2002).

5. A special case and a generalization

In the smooth function of means model, the parameter of interest $\gamma = g(\mu)$ is a smooth function of a k -dimensional vector mean $\mu = (\mu_1, \dots, \mu_k)' = E(X)$. Then μ is estimated by $\hat{\mu} = \bar{X} = n^{-1} \sum X_i$, and $\hat{\gamma} = g(\hat{\mu})$. This problem is a special case of the M -estimation framework with $\theta = \mu$ and $\psi(X, \theta) = X - \mu$. Then, $\hat{\sigma}^2 = \hat{g}'_\theta \hat{\Sigma} \hat{g}_\theta$ and

$$\hat{k}_3 = -n^{-1/2} \left[2n^{-1} \sum_{i=1}^n \{(X_i - \bar{X})' \hat{g}_\theta\}^3 + 3\hat{\delta}' \hat{g}_{\theta\theta} \hat{\delta} \right] / \hat{\sigma}^3,$$

where $\hat{\Sigma} = n^{-1} \sum (X_i - \bar{X})(X_i - \bar{X})'$, $\hat{\delta} = -\hat{\Sigma} \hat{g}_\theta$, $\hat{g}_{\theta\theta} = g_{\theta\theta}(\hat{\theta})$, and $g_{\theta\theta}(\theta) = \partial^2 g(\theta) / \partial \theta \partial \theta'$.

The results here are described in the framework of independent and identically distributed observations. However, only the assumption of independence is crucial. In many important M -estimation settings, the estimating function ψ changes for each observation, so that the contribution to the estimating function for observation X_i is written as $\psi_i(X_i, \theta)$, not as $\psi(X_i, \theta)$. For example, in regression problems, the function $\psi_i(X_i, \theta)$ depends on covariate values specific to observation X_i . In this more general context, the M -estimator $\hat{\theta}$ satisfies $\sum \psi_i(X_i, \hat{\theta}) = 0$, and all of the preceding results apply with $\psi_i(X_i, \theta)$ being used to replace $\psi(X_i, \theta)$ throughout. In the formula $V = A^{-1}\Sigma(A^{-1})'$, the components of the matrices $A = (A_{a/b})$ and $\Sigma = (\Sigma_{ab})$ are given by

$$A_{a/b} = E \left\{ n^{-1} \sum_{i=1}^n \psi_{i,a/b}(X_i, \theta) \right\}, \quad \Sigma_{ab} = E \left\{ \sum_{i=1}^n \psi_{i,a}(X_i, \theta) \psi_{i,b}(X_i, \theta) \right\},$$

where $\psi_i(X_i, \theta) = \{\psi_{i,1}(X_i, \theta), \dots, \psi_{i,k}(X_i, \theta)\}'$ and $\psi_{i,a/b}(X_i, \theta) = \partial \psi_{i,a}(X_i, \theta) / \partial \theta^b$.

6. Examples

The following three sets of figures pertain to three different situations. For each situation, the figure labeled (a) shows plots of s_γ^2 (dashed line) and s_τ^2 (solid line) versus γ , where $s_\tau = (\hat{\gamma}^\lambda - \gamma^\lambda) / (\lambda \hat{\sigma}_\gamma^2 \hat{\gamma}^{\lambda-1})$ is the studentized statistic based on the reparameterization $\tau = (\gamma^\lambda - 1) / \lambda$ with λ given by expression (5). The figure labeled (b) shows a Q-Q plot for 999 bootstrap values of s_γ . The figure labeled (c) shows a Q-Q plot for 999 bootstrap values of s_τ ; in the resampling process for s_τ , the value of λ is held fixed. In figures (b) and (c), the normal quantiles on the horizontal axis are adjusted for the bootstrap estimates of mean and standard deviation of s_γ and s_τ obtained from the 999 resamples.

Figures 1(a), 1(b), and 1(c) are based on the cd4 data analyzed by DiCiccio & Efron (1996). This dataset consists of a sample of size $n = 20$ drawn from a bivariate distribution. The interest parameter γ is the variance of the first component. Here, $\hat{\gamma} = 0.6242$, $\hat{\sigma}_\gamma = 0.7953$, and $\hat{k}_3 = -0.8452$; expression (5) yields $\lambda = 0.0111$, so τ is nearly the logarithmic transformation. The bootstrap estimates of the mean and standard deviation of s_γ are -0.5212 and 1.4741 ; the estimates for s_τ are -0.2994 and 1.214 .

Figures 2(a), 2(b), and 2(c) are also based on the cd4 data, but the interest parameter γ is the largest eigenvalue of the variance matrix of the underlying distribution. Here, $\hat{\gamma} = 1.6753$, $\hat{\sigma}_\gamma = 1.8222$, and $\hat{k}_3 = -0.4796$; expression (5) yields $\lambda = 0.3427$, so τ is close to the cube-root transformation. The bootstrap estimates of the mean and standard deviation for s_γ are -0.3935 and 1.2767 ; the estimates for s_τ are -0.2799 and 1.1547 .

Figures 3(a), 3(b), and 3(c) are based on the air-conditioning data failure ratio problem of Davison & Hinkley (1997, p. 217). This dataset consists of two samples of sizes $n_1 = 12$ and $n_2 = 24$ drawn from two populations having positive means μ_1 and μ_2 ; the parameter of interest is $\gamma = \mu_1 / \mu_2$. Here, $\hat{\gamma} = 0.5933$, $\hat{\sigma}_\gamma = 1.4216$, and $\hat{k}_3 = -1.1325$; expression (5) yields $\lambda = 0.0547$, so τ is close to the logarithmic transformation. The bootstrap estimates of the mean and standard deviation for s_γ are -0.1797 and 1.2799 ; the estimates for s_τ are 0.0656 and 1.2234 .

In each case, the magnitude of $\hat{k}_3$ is large, which indicates that both the profile empirical log likelihood for γ and the distribution of $s_\gamma = \sqrt{n}(\hat{\gamma} - \gamma) / \hat{\sigma}_\gamma$ are substantially skewed. Indeed, Figures 1(a), 2(a), and 3(a) show that the standard symmetric intervals

obtained from s_γ and s_τ will be quite different in shape; those obtained from s_τ will be closer in shape to empirical likelihood ratio intervals. Figures 1(b), 2(b), and 3(b) indicate that the distribution of s_γ is skewed; indeed, a normal approximation to the bootstrap distribution of s_γ is poor, even if that approximation is adjusted for mean and standard deviation. In contrast, Figures 1(c), 2(c), and 3(c) show that the reparameterization effectively eliminates the skewness. The standard normal approximation to the bootstrap distribution of s_τ is more accurate than it is to the bootstrap distribution of s_γ , and the normal approximation for s_τ is quite accurate if it is adjusted for mean and standard deviation.

The number of resamples required to estimate the mean and standard deviation of a test statistic is far fewer than the number of resamples required to estimate its entire bootstrap distribution. Consequently, the accuracy of the bootstrap- t method can be enjoyed with far less computation by using a normal approximation for s_τ with bootstrap estimates for mean and standard deviation.

Several authors, including Owen (1988), have pointed out that the χ_1^2 approximation to the distribution of w_γ can be inaccurate, so they have recommended that the distribution of w_γ be approximated by its bootstrap distribution. Bootstrap calibration for w_γ can be computationally very demanding. A more computationally tractable alternative is to bootstrap calibrate s_τ^2 ; since s_τ^2 and w_γ are close, it is likely that bootstrapping s_τ^2 will produce similar results to bootstrapping w_γ .

References

- Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.
- DiCiccio, T. J. & Efron, B. (1996). Bootstrap confidence intervals (with Discussion). *Statist. Sci.* **11**, 189-228.
- DiCiccio, T. J. & Monti, A. C. (2001). Accurate confidence limits for scalar functions of vector M -estimands. *Biometrika*, to appear.
- DiCiccio, T. J. & Monti, A. C. (2001). Approximations to the profile empirical likelihood function for a scalar parameter in the context of M -estimation. *Biometrika* **88**, 337-352.
- DiCiccio, T. J. & Tibshirani, R. (1987). Bootstrap confidence intervals and bootstrap approximations. *J. Amer. Statist. Assoc.* **82**, 163-170.
- Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.
- Konishi, S. (1991). Normalizing transformations and bootstrap confidence intervals. *Ann. Statist.* **19**, 2209-2225.
- Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237-249. Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90-120.
- Sprott, D. A. (1980). Maximum likelihood in small samples: Estimation in the presence of nuisance parameters. *Biometrika* **67**, 515-523.





