A discussion of 'A selective review on calibration information from similar studies'

Jiahua Chen

Published online: 31 Aug 2022.

Submit your article to this journal ↗

Article views: 36

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

SHORT COMMUNICATION

A discussion of 'A selective review on calibration information from similar studies'

Jiahua Chen 🔸

Department of Statistics, University of British Columbia, Vancouver, BC, Canada

Being a long-time friend of Dr. Qin and served as a supervisor of Drs. Li and Liu, I am as proud as authors of the richness of the content as well as the broadness of this paper. It helps me to play catch up and shames me to work hard rather than hardly work.

As a discussant, I wish to come up with some additional insight on this research topic but this is deemed a very difficult task. I should congratulate the authors for covering a vast territory and leave no room for that. Instead, I raise two not so important technical issues which might be of interest to some fellow researchers.

1. Combination of the empirical likelihood and estimating function

Likelihood-based approaches have many easy-to-utilize properties for statistical inference. Under regularity conditions on a parametric statistical model and with a set of independent and identically distributed observations, the maximum likelihood estimator of the model parameter is asymptotically normal and has the lowest possible asymptotic covariance matrix, the likelihood ratio test of certain hypotheses generally has superior power, and the test statistic has a convenient chi-squared limiting distribution. These statements are granted, however, by implicitly assuming the users can come up with a defensible parametric model for the data they collected. In many industries, users are reluctant to put forward a parametric model for various reasons. Can there be a nonparametric likelihood that possesses most of the useful properties of the parametric likelihood? The groundbreaking work of Owen (1988, 1990) on the empirical likelihood (EL) confirms this possibility. His initial developments focus on inferences on parameters that can be regarded as population means or regression parameters. Qin and Lawless (1995) greatly empower the EL by introducing estimating equations (EE) under this umbrella. They form the basis of a new branch of mathematical statistics.

Presented slightly differently from the paper being discussed, suppose the data generating distribution F is a member of a nonparametric collection of distributions denoted by $\mathcal{F}$. It is further known that

$$\mathbb{E}\{g(X; \theta)\} = 0$$

for some value θ, with the expectation calculated when X has distribution F and the function $g(X; \theta)$ is known. Such a model setup uses θ to summarize certain aspects of the population distribution and focuses on the statistical inference regarding θ. The model $\mathcal{F}$ does not possess other parametric structures. Amazingly, Qin and Lawless (1995) construct the profile EL for θ and find its nonparametric maximum likelihood estimator (MLE) $\hat{\theta}$ is asymptotic normal with the lowest possible asymptotic variance under some conditions mostly on $g(X; \theta)$.

Unlike many asymptotic results in some papers, the conditions here are simple and easy to understand. Two major conditions include that $g(X; \theta)$ being smooth in θ and that $g(X; \theta)$ and its derivatives with respect to θ having finite moments up to order 3. In other words, the asymptotic conclusions regarding nonparametric MLE are valid under most general conditions and therefore widely applicable. Users are practically free from any model misspecification risk.

There is an important case, population quantile, that their results do not directly apply. An estimating function for the univariate population median, for instance, is given by

$$g(X; \theta) = 1 - 2\,\mathbf{1}(X \leq \theta).$$

Suppose this is the only estimating function regarding θ in the model, we do not need a general theorem regarding the large sample properties of its nonparametric MLE under EL. We can directly work on them. Such luck does not extend generally. Some researchers studied the profile EL with a non-smooth estimating function (Molanes Lopez et al., 2009). Their conclusions,

however, generally lose the simplicity and the beauty of these given in Qin and Lawless (1995). Can we come up with a non-scary set of conditions for non-smooth $g(X; \theta)$ under which the asymptotic results of Qin and Lawless (1995) hold? This question has been in my mind for some time.

Another generally overlooked deficiency, however minor, is the locality of some asymptotic results. To make it clear, let us review a key technical step in the proof of consistency and asymptotic normality of the nonparametric MLE $\hat{\theta}$. Denote the true parameter value by θ^*, the sample size n and the log profile empirical likelihood $\ell_n(\theta)$. One proves that $\ell_n(\theta^* + n^{-1/3}\mathbf{a}) < \ell_n(\theta^*)$ with probability approaching 1 as $n \to \infty$ uniformly with respect to unit vector $\mathbf{a}$. Hence, there must be a local maximum within $n^{-1/3}$ distance of the true parameter value θ^* under some smoothness condition on $g(X; \theta)$. It is this local nonparametric MLE, to name it so, that has all the claimed nice asymptotic properties. This line of approach is widely attributed to Cramér (2016) and can be found in Shao (2003).

At the same time, the consistency of the global MLE under a parametric model with minimum conditions is given by Wald (1949). By global, I require $\hat{\theta}$ to be the maximum point of $\ell_n(\theta)$ over the whole parameter space Θ of θ. The same result must be true for the nonparametric MLE of Qin and Lawless (1995), only if we can lay out a set of easily verifiable and reasonable conditions on $g(X; \theta)$.

Let me deliberate it further. Following the general practice, let $x_1, \ldots, x_n$ be a set of observations on independent and identically distributed random variables with common $F \in \mathcal{F}$. They may also be regarded as random variables wherever appropriate. I will also use X for a random variable with distribution F. Suppose $\mathcal{F}$ is parametric containing density functions $f(x; \theta)$ indexed by $\theta \in \Theta$. The Jensen inequality is valid almost universally:

$$\mathbb{E}\{\log f(X; \theta)\} \leq \mathbb{E}\{\log f(X; \theta^*)\}$$

when the true parameter value is θ^*. This leads to

$$\sum_{i=1}^{n} \log f(x_i; \theta) \leq \sum_{i=1}^{n} \log f(x_i; \theta^*) \qquad (1)$$

almost surely for any *fixed* $\theta \neq \theta^*$. The consistency proof of Wald (1949). is to show that this inequality holds uniformly outside any neighbourhood of θ^*. Do we have a similar inequality for the profile EL?

Given any θ, let $\lambda = \lambda(\theta)$ be the Lagrange multiplier that is the by-product of solving the constrained optimization problem regarding profile EL. It is well known that λ satisfies (or with probability approaching 1)

$$\sum_{i=1}^{n} \frac{g(x_i; \theta)}{1 + \lambda^{\top} g(x_i; \theta)} = 0. \qquad (2)$$

The profile log-likelihood function, after omitting an additive constant $n \log n$, is given by

$$\ell_n(\theta) = -\sum_{i=1}^{n} \log\{1 + \lambda^{\top} g(x_i; \theta)\}.$$

Let us study this expression from a different angle following Chen et al. (2008). Given θ, we define a function of γ as follows:

$$\psi(\gamma) = -\sum_{i=1}^{n} \log\{1 + \gamma^{\top} g(x_i; \theta)\}.$$

It is seen that $\psi(\gamma)$ is convex because its second derivative matrix is positive definite. There is a negligible chance to be semi-positive definite. Hence, $\psi(\gamma)$ is convex and attains the minimum at $\gamma = \lambda = \lambda(\theta)$ given by (2). Consequently, we have

$$\psi(\gamma) \geq \psi(\lambda) = \ell_n(\theta).$$

Let θ be such that $\mathbb{E}\{g(X; \theta)\} = \delta \neq 0$. When $g(X; \theta)$ has finite third absolute moment, following Owen (2001), we have

$$\max_{i} \|g(x_i; \theta)\| = o_p(n^{1/3}).$$

Let

$$\gamma = n^{-2/3}\bar{g}_n = n^{-5/3} \sum_{i=1}^{n} g(x_i; \theta)$$

so that

$$\max_{i} \|\gamma^{\top} g(x_i; \theta)\| = o_p(n^{-1/3}) = o_p(1).$$

Note that $\log(1 + t) = t + O(t^2)$ when $t \to 0$ and keep in mind about the uniformity of the above claim in i. Applying this fact with the current γ, we find with a probability approaching 1,

$$\psi(\gamma) = -n^{1/3}\|\bar{g}_n\|^2 \to -\infty$$

at rate $n^{1/3}$ as $n \to \infty$. Hence, we have shown

$$\ell_n(\theta) = \psi(\lambda) \leq \psi(\gamma) \to -\infty.$$

From Qin and Lawless (1995), we already know that $2\ell_n(\theta^*)$ has a chisquare limiting distribution. Hence, we have $\ell_n(\theta^*) = O_p(1)$ and

$$\ell_n(\theta) < \ell_n(\theta^*) \qquad (3)$$

in probability. I leave it as a future project to proving the ultimate consistency of the global MLE.

2. The efficiency of nonparametric MLE

I am enlightened by the optimality of the nonparametric MLE under the model specified by an estimating function. The nonparametric model may be more precisely presented as

$$\mathcal{F} = \{F : \mathbb{E}\{g(X; \theta)\} = 0 \text{ for some } \theta \in \Theta\} \quad (4)$$

with some smoothness and moment requirements omitted. Let there be a parametric family containing distributions whose density functions are given by

$$\mathcal{P} = \{f(x; \theta) : \theta \in \Theta\}.$$

Assume the true distribution F^* of X is a member of $\mathcal{P}$. Let the true parameter value be θ^*. Given these, F^* is also a member of

$$\mathcal{H} = \{h(x; \eta, \theta) = C(\gamma, \theta) \exp\{\gamma^\top g(x; \theta)\} f(x; \theta)$$
$$: (\gamma, \theta) \in \Gamma \times \Theta\} \quad (5)$$

for some normalization constant $C(\gamma, \theta)$. I use γ instead of η in order to use Γ for its parameter space. The true distribution F^* under model $\mathcal{H}$ has density function $h(x; 0, \theta^*)$. Because $\mathcal{H}$ is a regular parametric family, its parametric MLE of (η, θ) is asymptotically efficient: they have the lowest asymptotic variance matrix. The paper points out that the asymptotic variances of the parametric and nonparametric MLE of θ are the same. Consequently, the nonparametric MLE of θ has attained the lowest possible asymptotic variance.

There is a hidden agenda behind the above discussion. One cannot attain a lower asymptotic variance matrix for estimating (η, θ) jointly. Still, how should we conclude the nonexistence of a more efficient estimator of θ from here? I cannot come up with a quick justification. In fact, Qin and Lawless (1995) has solid proof that the nonparametric MLE attains the lowest possible asymptotic variance of the estimator based on the optimal estimating function derived from $g(x; \theta)$.

Incidentally, I did some calculations and found the gradient

$$\nabla_\theta \log\{h(x; \theta^*, 0)\} = \frac{f'(x; \theta^*)}{f(x; \theta^*)}.$$

This is in contradiction of $\nabla_\theta \log\{h(x; \theta^*, 0)\} = 0$ given in the paper in discussion. Interestingly, may we replace $f(x; \theta)$ by the true density $f^*(x)$ of F^*, whatever the σ-finite algebra? If so, we still have a well defined extended model $\mathcal{H}$. Now we have $\nabla_\theta \log\{h(x; \theta^*, 0)\} = 0$ which further implies $\mathbb{E}\{\nabla_\theta h(X; \theta^*, 0)\} = 0$ so that the Fisher

information matrix is block orthogonal. Consequently, it also addresses my previous doubt about the gap between the joint efficiency and marginal efficiency. However, I am not confident in the validity of my suggested choice of $f^*(x)$.

3. Wrap up

Many of my recent research problems came from reading the papers of the first author of this discussion paper. Naturally, I have many join work with other two authors. If history is a reflection of the future, I should greatly benefit from studying this discussion paper. I urge young statisticians with similar research to follow my example.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Jiahua Chen 🔟 http://orcid.org/0000-0001-8064-4444

References

Chen, J., Variyath, A. M., & Abraham, B. (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, *17*(2), 426–443. https://doi.org/10.1198/106186008X321068

Cramér, H. (2016). Mathematical methods of statistics. In *Mathematical methods of statistics (PMS-9)* (Vol. 9). Princeton University Press.

Molanes Lopez, E. M., Keilegom, I. V., & Veraverbeke, N. (2009). Empirical likelihood for non-smooth criterion functions. *Scandinavian Journal of Statistics*, *36*(3), 413–432. https://doi.org/10.1111/sjos.2009.36.issue-3

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*(2), 237–249. https://doi.org/10.1093/biomet/75.2.237

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, *18*(1), 90–120. https://doi.org/10.1214/aos/1176347494

Owen, A. B. (2001). *Empirical likelihood*. Chapman & Hall/CRC.

Qin, J., & Lawless, J. (1995). Estimating equations, empirical likelihood and constraints on parameters. *Canadian Journal of Statistics*, *23*(2), 145–159. https://doi.org/10.2307/3315441

Shao, J. (2003). *Mathematical statistics*. Springer Science & Business Media.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, *20*(4), 595–601. https://doi.org/10.1214/aoms/1177729952