Higher-order properties of Bayesian empirical likelihood

Xiaolong Zhong and Malay Ghosh

Department of Statistics, University of Florida
32611, Florida, USA
e-mail: zhongxl@ufl.edu*;* ghoshm@stat.ufl.edu

Abstract: Empirical likelihood serves as a good nonparametric alternative to the traditional parametric likelihood. The former involves much less assumptions than the latter, but very often gets the same asymptotic inferential efficiency. While empirical likelihood has been studied quite extensively in the frequentist literature, the corresponding Bayesian literature is somewhat sparse. Bayesian methods hold promise, however, especially with the availability of historical information, which often can be used successfully for the construction of priors. In addition, Bayesian methods very often overcome the curse of dimensionality by providing suitable dimension reduction through judicious use of priors and analyzing data with the resultant posteriors. In this paper, we provide asymptotic expansion of posteriors for a very general class of priors along with the empirical likelihood and its variations, such as the exponentially tilted empirical likelihood and the Cressie–Read version of the empirical likelihood. Other than obtaining the celebrated Bernstein–von Mises theorem as a special case, our approach also aids in finding non-subjective priors based on empirical likelihood and its variations as mentioned above.

Keywords and phrases: Bernstein-von Mises theorem, Cressie–Read divergence, exponentially tilted empirical likelihood.

Received July 2015.

1. Introduction

Empirical likelihood, over the years, has become a very popular topic of statistical research. The name was coined by Owen in his classic 1986 paper, although similar ideas are found even earlier in the works of [11], [19], [16] and others. The main advantage of empirical likelihood is that it involves fewer assumptions than a regular likelihood, and yet shares the same asymptotic properties of the latter.

Research in this area has primarily been frequentist with a long list of important theoretical developments accompanied by a large number of applications. To our knowledge, the first Bayesian work in this general area appeared in the article of [13] followed by some related work in [17, 18], the latter introducing the concept of "exponentially tilted empirical likelihood". [13] suggested using empirical likelihood as a substitute for the usual likelihood and carrying out Bayesian analysis in the usual way.

Baggerly [1] viewed empirical likelihood as a method of assigning probabilities to a n-cell contingency table in order to minimize a goodness-of-fit criterion. He selected Cressie–Read power divergence statistics as one such criterion for

construction of confidence regions in a number of situations and pointed out also how the usual empirical likelihood, exponentially tilted empirical likelihood and others could be viewed as special cases of the Cressie–Read criterion by appropriate choice of the the power parameter. This was also discussed in [15] who pointed out that all members of the Cressie–Read family led to "empirical divergence analogues of the empirical likelihood in which asymptotic χ^2 calibration held for the mean".

The objective of this article is to provide an asymptotic expansion of the posterior distribution based on empirical likelihood and its variations under certain regularity conditions and a mean constraint. The work is inspired by the work of [5] who provided a somewhat different expansion subject to a mean constraint. Unlike [4, 5], our result is based on the derivatives of the pseudo likelihood with respect to the parameter of interest evaluated at the maximum empirical likelihood estimator, and a rigorous expansion is provided with particular attention to the remainder terms. Moreover, we consider a general estimating equation which includes the mean example of [5] as a special case. The need for different pseudo-likelihoods for statistical inference is felt all the more in these days, especially for the analysis of high-dimensional data, where the usual likelihood based analysis is hard to perform, These alternative likelihoods are equally valuable for approximate Bayesian computations, a topic which has only recently surfaced in the statistics literature (see e.g. [3]).

Asymptotic expansion of the posterior based on a regular likelihood was given earlier in [12], and later in [8]. We follow their approach with many necessary modifications in view of the fact that any meaningful prior needs to have support in a data-driven compact set which grows with number of observations. As a special case of our result, we get the celebrated Bernstein–von Mises theorem. The latter was mentioned in [13] for the special case of empirical likelihood, but here we provide a rigorous derivation with the needed regularity conditions in a general framework. The asymptotic expansion can also be used in providing asymptotic expansions of the posterior moments, quantiles and other quantities of interest. Moreover, we utilize this asymptotic expansion to find some moment matching priors, earlier given in [9] based on the regular likelihood. In constrast to [9], the moment matching prior does not depend on the expectation of the derivatives of the log-likelihood function, but depends instead on the second and third central moments of the unbiased estimating function, say $g(X, \theta)$. In the particular case, $g(X, \theta) = X - \theta$, the prior depends only on knowledge about the second and third central moments of the distribution, and does not require specification of a full likelihood. The moment matching priors differ also from the reference priors as introduced in [2]. The latter is an analogue of Jeffreys' prior under most circumstances, with the Godambe information matrix [10] replacing the Fisher information matrix.

2. Basic settings

Suppose $X_1, \ldots, X_n$ are independent and identically distributed random vectors satisfying $E\{g(X_1, \theta)\} = 0$, where $\theta \in \mathbb{R}$. In this context, [14], formulated

empirical likelihood as a nonparametric likelihood of the form $\prod_{i=1}^{n} w_i(\theta)$, where w_i is the probability mass assigned to $X_i (i = 1, \ldots, n)$ satisfying the constraints

$$\begin{cases} w_i > 0, \text{ for all i;} \\ \sum_{i=1}^{n} w_i = 1; \\ \sum_{i=1}^{n} w_i g(X_i, \theta) = 0. \end{cases} \tag{2.1}$$

The target is to maximize $\prod_{i=1}^{n} w_i$ or equivalently $\sum_{i=1}^{n} \log w_i$ with respect to $w_1, \ldots, w_n$ subject to the constraints given in Eq. (2.1). Applying the Lagrange multiplier method, the solution turns out to be

$$\hat{w}_i^{\text{EL}}(\theta) = \frac{1}{n\{1 + \nu g(X_i, \theta)\}}, \; i = 1, \ldots, n, \tag{2.2}$$

where ν, the Lagrange multiplier satisfies

$$\sum_{i=1}^{n} \frac{g(X_i, \theta)}{1 + \nu g(X_i, \theta)} = 0. \tag{2.3}$$

It may be noted that in [4, 5], $g(X_i, \theta) = X_i - \theta, \; i = 1, \ldots, n$.

Closely related to the empirical likelihood is the exponentially tilted empirical likelihood where the objective is to maximize the Shannon entropy $-\sum_{i=1}^{n} w_i \log w_i$ with the same constraints in Eq. (2.1). The resulting solution is

$$\hat{w}_i^{\text{ET}}(\theta) = \frac{\exp\{-\nu g(X_i, \theta)\}}{\sum_{j=1}^{n} \exp\{-\nu g(X_i, \theta)\}},$$

where ν, the Lagrange multiplier, satisfies

$$\sum_{i=1}^{n} \exp\{-\nu g(X_i, \theta)\} g(X_i, \theta) = 0. \tag{2.4}$$

The exponentially tilted empirical likelihood is related to Kullback-Leibler divergence between two empirical distributions, one with weights w_i assigned to the n sample points, and the other with uniform weights $1/n$ assigned to the sample points.

The general Cressie–Read divergence criterion given by

$$\text{CR}(\lambda) = \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^{n} \left\{ (nw_i)^{-\lambda} - 1 \right\}.$$

We focus on the cases $\lambda \geq 0$ and $\lambda \leq -1$, because in these cases, $\text{CR}(\lambda)$ is a convex function of the $w_i (i = 1, \ldots, n)$, and hence the minimization problem will produce a unique solution. The limiting cases $\lambda \to 0$ and $\lambda \to -1$ correspond to the usual empirical likelihood and the exponentially tilted empirical likelihood as defined earlier.

For convex CR(λ), its minimum will be attained in the compact set H_n determined by data. The Lagrange multiplication method now gives the weights

$$\hat{w}_i^{\mathrm{CR}}(\theta) = \frac{1}{n} \left\{ \mu + \nu g(X_i, \theta) \right\}^{-1/(\lambda+1)}, \, i = 1, \ldots, n, \qquad (2.5)$$

where we abbreviate $\mu(\theta)$ as μ and $\nu(\theta)$ as ν, which satisfy

$$\begin{cases} \sum_{i=1}^n \left\{ \mu + \nu g(X_i, \theta) \right\}^{-1/(\lambda+1)} = n, \\ \sum_{i=1}^n \left\{ \mu + \nu g(X_i, \theta) \right\}^{-1/(\lambda+1)} X_i = 0. \end{cases} \qquad (2.6)$$

We now introduce the posterior based on an empirical likelihood. The basic idea was first introduced by [13] with several numerical examples. The intuition relies on close relationship between the empirical likelihood and the empirical distribution. [15] formulated the two concepts under the same optimization framework, that is, they shared the same objective function, but the former was solved under parametric constraints, while the latter was not. Considering this similarity, we can use the empirical likelihood as a valid distribution parameterized by some inferential target. Within the Bayesian paradigm, writing $\hat{w}_i(\theta)$ as generic notation for either $\hat{w}_i^{\mathrm{EL}}$, $\hat{w}_i^{\mathrm{ET}}$ or $\hat{w}_i^{\mathrm{CR}}$, and a prior with probability density function $\rho(\theta)$, with support in H_n, the profile (pseudo) posterior is

$$\pi(\theta \mid X_1, \ldots, X_n) = \frac{\prod_{i=1}^n \hat{w}_i(\theta) \rho(\theta)}{\int_{H_n} \prod_{i=1}^n \hat{w}_i(\theta) \rho(\theta) \, \mathrm{d}\theta}. \qquad (2.7)$$

The main objective of this paper is to provide an asymptotic expansion of $\pi(\theta \mid X_1, \ldots, X_n)$. This will include in particular the Bernstein–von Mises theorem. Towards our main result, we develop a few necessary lemmas in the next section. Some of these lemmas are also of independent interest as they point out some interesting features pertaining to empirical likelihood.

3. Lemmas

We first point out the natural domain of θ in empirical likelihood settings. In practice, some values of θ will result in an empty feasible set under constraints Eq. (2.1). The set of θ values which guarantees a non-empty feasible set, and thus a solution of the optimization problem, constitutes a natural domain of the empirical likelihood. One may question whether the size of the natural domain is large enough to contain the true value. The following lemma alleviates this worry.

Lemma 1. *Assume $g(\cdot, \cdot)$ is a continuous function, then the natural domain defined by the constraints Eq. (2.1) is a compact set and is nondecreasing with respect to the sample size n.*

Proof. By the third constraint of Eq. (2.1), θ is a continuous function of $w_1, w_2, \ldots, w_n$, but w_i are defined on a simplex which is a compact set due

to the first constraint of Eq. (2.1). We may recall that a continuous function maps compact sets to compact sets. Hence, θ is naturally defined on a compact set denoted by H_n.

If all the $g(X_i, \theta), i = 1, \ldots, n$ are non-positive or all are non-negative, then the constraints Eq. (2.1) are violated and $H_n = \emptyset$. Hence, we define the domain as

$$
\begin{aligned}
H_n &= \left(\left[\bigcap_{i=1}^{n} \{ g(X_i, \theta) \geq 0 \} \right] \bigcup \left[\bigcap_{i=1}^{n} \{ g(X_i, \theta) \leq 0 \} \right] \right)^c \\
&= \left[\bigcup_{i=1}^{n} \{ g(X_i, \theta) \geq 0 \}^c \right] \bigcap \left[\bigcup_{i=1}^{n} \{ g(X_i, \theta) \leq 0 \}^c \right].
\end{aligned}
$$

As n increases, both $\{ [\bigcup_{i=1}^{n} \{ g(X_i, \theta) \geq 0 \}^c] \}$ and $\{ [\bigcup_{i=1}^{n} \{ g(X_i, \theta) \leq 0 \}^c] \}$ will increase, and so will their intersection H_n. □

Although, intuitively we expect the empirical likelihood to behave as the true likelihood, we need some theoretical support to show that the former enjoys some of the basic properties of the latter. In particular, we need to verify that ν and μ are smooth functions of θ and the (pseudo) Fisher Information based on the empirical likelihood is positive.

We first establish the positiveness of the Fisher information. We consider the three cases separately to introduce more transparency and continuity in our approach.

Our first lemma shows that the Lagrange multipliers $\nu(\theta)$ and $\mu(\theta)$ are both smooth functions of θ, under the following mild assumptions,

Assumption 1. For any θ in natural domain H_n, and $n \geq 3$,

$$
\text{pr} \{ g(X_i, \theta) = 0, \ i = 1, \ldots, n \} = 0.
$$

And

Assumption 2. $g(x, \theta)$ is a continuous multivariate function with continuous derivatives in θ.

Lemma 2. *Under Assumptions 1 and 2, for the empirical likelihood, exponentially tilted empirical likelihood and Cressie–Read empirical likelihood (λ), the Lagrange multipliers $\nu(\theta)$ and $\mu(\theta)$ are smooth functions of θ.*

Proof. We first consider the empirical likelihood and observe that, $\nu(\theta)$ is a implicit function of θ in view of (2.3). Further

$$
\frac{\partial}{\partial \nu} \sum_{i=1}^{n} \frac{g(X_i, \theta)}{1 + \nu g(X_i, \theta)} = - \sum_{i=1}^{n} \frac{g^2(X_i, \theta)}{\{ 1 + \nu g(X_i, \theta) \}^2} < 0,
$$

so that by the implicit function theorem, ν is differentiable in θ. Moreover, differentiating both sides of Eq. (2.3) with respect to θ, one gets

$$
0 = \sum_{i=1}^{n} \frac{1}{1 + \nu g(X_i, \theta)} \frac{\mathrm{d} g(X_i, \theta)}{\mathrm{d} \theta} - \sum_{i=1}^{n} \frac{\nu g(X_i, \theta)}{\{ 1 + \nu g(X_i, \theta) \}^2} \frac{\mathrm{d} g(X_i, \theta)}{\mathrm{d} \theta}
$$

$$-\sum_{i=1}^{n} \frac{g^2\left(X_i, \theta\right)}{\left\{1+\nu g\left(X_i, \theta\right)\right\}^2} \frac{\mathrm{d}\nu}{\mathrm{d}\theta},$$

which on simplification leads to

$$\frac{\mathrm{d}\nu}{\mathrm{d}\theta} = -\frac{\sum_{i=1}^{n}\left\{1+\nu g\left(X_i, \theta\right)\right\}^{-2} \mathrm{d}g\left(X_i, \theta\right)/\mathrm{d}\theta}{\sum_{i=1}^{n}\left\{1+\nu g\left(X_i, \theta\right)\right\}^{-2} g^2\left(X_i, \theta\right)}, \tag{3.1}$$

Next, for exponentially tilted empirical likelihood, in view of Eq. (2.4) and the relation

$$\frac{\mathrm{d}}{\mathrm{d}\nu}\left[\sum_{i=1}^{n} \exp\left\{-\nu g\left(X_i, \theta\right)\right\} g\left(X_i, \theta\right)\right]$$

$$= -\sum_{i=1}^{n} \exp\left\{-\nu g\left(X_i, \theta\right)\right\} g^2\left(X_i, \theta\right) < 0,$$

once again, the implicit function theorem guarantees the differentiability of ν in θ. Further, differentiating both sides of Eq. (2.4) with respect to θ, one gets

$$\frac{\mathrm{d}\nu}{\mathrm{d}\theta} = \frac{\sum_{i=1}^{n} \exp\left\{-\nu\left(\theta\right) g\left(X_i, \theta\right)\right\}\left\{\mathrm{d}g\left(X_i, \theta\right)/\mathrm{d}\theta\right\}\left\{1-\nu g\left(X_i, \theta\right)\right\}}{\sum_{i=1}^{n} \exp\left\{-\nu\left(\theta\right) g\left(X_i, \theta\right)\right\} g^2\left(X_i, \theta\right)}. \tag{3.2}$$

A similar conclusion is achieved for $\nu(\theta)$ and $\mu(\theta)$ defined in Eq. (2.6) in connection with CR(λ). Specifically, defining

$$\begin{cases} F_1 = \sum_{i=1}^{n}\left\{\mu+\nu g\left(X_i, \theta\right)\right\}^{-1/(\lambda+1)} - n, \\ F_2 = \sum_{i=1}^{n}\left\{\mu+\nu g\left(X_i, \theta\right)\right\}^{-1/(\lambda+1)} g\left(X_i, \theta\right), \end{cases}$$

it follows that,

$$\frac{\partial\left(F_1, F_2\right)}{\partial\left(\mu, \nu\right)} = -\frac{1}{\lambda+1}\begin{pmatrix} \sum_{i=1}^{n} q_i & \sum_{i=1}^{n} q_i g\left(X_i, \theta\right) \\ \sum_{i=1}^{n} q_i g\left(X_i, \theta\right) & \sum_{i=1}^{n} q_i g\left(X_i, \theta\right)^2 \end{pmatrix},$$

where $q_i = \left\{\mu+\nu g(X_i, \theta)\right\}^{-1/(\lambda+1)-1}$. Then the determinant of Jacobian is

$$\begin{aligned}
\det \frac{\partial\left(F_1, F_2\right)}{\partial\left(\mu, \nu\right)} &= \left(\frac{1}{\lambda+1}\right)^2\left[\sum_{i=1}^{n} q_i \sum_{i=1}^{n} q_i g\left(X_i, \theta\right)^2 - \left\{\sum_{i=1}^{n} q_i g\left(X_i, \theta\right)\right\}^2\right] \\
&= \left(\frac{1}{\lambda+1}\right)^2\left(\sum_{i=1}^{n} q_i\right)^2\left[\sum_{i=1}^{n} \frac{q_i}{\sum_{j=1}^{n} q_j} g\left(X_i, \theta\right)^2\right. \\
&\quad \left. -\left\{\sum_{i=1}^{n} \frac{q_i}{\sum_{j=1}^{n} q_j} g\left(X_i, \theta\right)\right\}^2\right] > 0.
\end{aligned}$$

Again, by implicit function theorem, one gets differentiability of $\mu(\theta)$ and $\nu(\theta)$ with respect to θ, and

$$\begin{pmatrix} \mu' \\ \nu' \end{pmatrix} = \left(\frac{\partial\left(F_1, F_2\right)}{\partial\left(\mu, \nu\right)}\right)^{-1}\begin{pmatrix} \partial F_1/\partial\theta \\ \partial F_2/\partial\theta \end{pmatrix}$$

$$= \left(-\frac{1}{\lambda+1} \right) (\lambda+1)^2 \frac{1}{\sum_{i=1}^n q_i \sum_{i=1}^n q_i g\left(X_i,\theta\right)^2 - \left\{ \sum_{i=1}^n q_i g\left(X_i,\theta\right) \right\}^2}$$

$$\times \begin{pmatrix} \sum_{i=1}^n q_i g\left(X_i,\theta\right)^2 & -\sum_{i=1}^n q_i g\left(X_i,\theta\right) \\ -\sum_{i=1}^n q_i g\left(X_i,\theta\right) & \sum_{i=1}^n q_i \end{pmatrix}$$

$$\times \left((\lambda+1)^{-1} \sum_{i=1}^n q_i \nu \frac{\mathrm{d} g\left(X_i,\theta\right)}{\mathrm{d}\theta}, -(\lambda+1) \sum_{i=1}^n q_i \nu g\left(X_i,\theta\right) \frac{\mathrm{d} g\left(X_i,\theta\right)}{\mathrm{d}\theta} \right.$$

$$\left. + \sum_{i=1}^n \left\{ \mu + \nu g\left(X_i,\theta\right) \right\}^{-1/(\lambda+1)} \frac{\mathrm{d} g\left(X_i,\theta\right)}{\mathrm{d}\theta} \right)^T. \tag{3.3}$$

$\square$

The next result shows that all the derivatives of the Lagrange multipliers $\nu(\theta)$ and $\mu(\theta)$ are smooth functions of $\theta \in H_n$. We provide a unified proof for all three cases where we utilize the previous lemma. with an assumption stronger than Assumption 2,

Assumption 3. $g(x,\theta)$ is a multivariate continuous function and $(K+4)$th-order differentiable in θ.

Lemma 3. *Under Assumptions 1 and 3, all orders of derivatives of $\nu(\theta)$ and $\mu(\theta)$ are smooth functions of θ for $\theta \in H_n$.*

Proof. The result is proved by induction. We have seen already in Lemma 2, that the first derivatives of $\nu'(\theta)$ and $\mu'(\theta)$ are smooth functions of θ. Suppose the result holds for all kth derivatives of $\nu(\theta)$ and $\mu(\theta)$ for $k = 1, \ldots, K$. Then writing

$$\frac{\mathrm{d}^k \nu}{\mathrm{d}\theta^k} = h_k \left\{ \nu\left(\theta\right), \theta \right\}, k = 1, \ldots, K,$$

$$\frac{\mathrm{d}^{k+1} \nu}{\mathrm{d}^{k+1}\theta} = \frac{\partial h_k}{\partial \nu} \frac{\mathrm{d}\nu}{\mathrm{d}\theta} + \frac{\partial h_k}{\partial \theta}$$

which is also a smooth function of θ by the induction hypothesis and Lemma 1. A similar proof works for $\mu(\theta)$. $\square$

We know that when the number of constraints and dimension of the parameters are the same, the corresponding empirical likelihood is maximized at $\theta = \tilde{\theta}$, the M-estimator of θ based on $\sum_{i=1}^n g(X_i,\theta) = 0$. Thus, $\nu(\tilde{\theta}) = 0$ and $\mu(\tilde{\theta}) = 1$. We next show that $\tilde{l}(\theta)$ has a negative second order derivative when evaluated at $\tilde{\theta}$.

Lemma 4. *Under Assumptions 1 and 2, $\mathrm{d}^2 \tilde{l}(\tilde{\theta}) / \mathrm{d}\theta^2 < 0$ where $\tilde{l}(\theta) = n^{-1} \sum_{i=1}^n \log \hat{w}_i(\theta)$ where $\hat{w}_i$ is either $\hat{w}_i^{\mathrm{EL}}$, $\hat{w}_i^{\mathrm{ET}}$ or $\hat{w}_i^{\mathrm{CR}}$ $(i = 1, \ldots, n)$.*

Proof. We begin with $\tilde{l}(\theta) = n^{-1} \sum_{i=1}^n \log \hat{w}_i^{\mathrm{EL}}(\theta) = -\sum_{i=1}^n \log\{1 + \nu(X_i - \theta)\} - \log n$. Hence by Eq. (2.2), Eq. (2.3) and Eq. (3.1),

$$\frac{\mathrm{d}\tilde{l}(\theta)}{\mathrm{d}\theta} = \frac{1}{n} \nu(\theta) \sum_{i=1}^n \frac{1}{1 + \nu g\left(X_i\theta\right)} \frac{\mathrm{d} g\left(X_i,\theta\right)}{\mathrm{d}\theta} - \frac{1}{n} \sum_{i=1}^n \frac{g\left(X_i\theta\right)}{1 + \nu g\left(X_i\theta\right)} \frac{\mathrm{d}\nu}{\mathrm{d}\theta}$$

$$= \frac{1}{n}\nu\left(\theta\right)\sum_{i=1}^{n}\frac{1}{1+\nu g\left(X_i\theta\right)}\frac{\mathrm{d}g\left(X_i,\theta\right)}{\mathrm{d}\theta}.$$

Thus

$$\left.\frac{\mathrm{d}^2\tilde{l}\left(\theta\right)}{\mathrm{d}\theta^2}\right|_{\theta=\tilde{\theta}}=-\frac{\left\{\sum_{i=1}^{n}\mathrm{d}g\left(X_i,\tilde{\theta}\right)/\mathrm{d}\theta\right\}^2}{n\sum_{i=1}^{n}g^2\left(X_i,\tilde{\theta}\right)}<0.$$

Next we consider $\tilde{l}(\theta) = n^{-1}\sum_{i=1}^{n}\log\hat{w}_i^{\mathrm{ET}}(\theta) = -\nu n^{-1}\sum_{i=1}^{n}g(X_i,\theta) - \log\sum_{i=1}^{n}\exp\{-\nu g(X_i,\theta)\}$. Then

$$
\begin{aligned}
\frac{\mathrm{d}\tilde{l}\left(\theta\right)}{\mathrm{d}\theta} &= -\frac{\mathrm{d}\nu}{\mathrm{d}\theta}\frac{1}{n}\sum_{i=1}^{n}g\left(X_i,\theta\right) \\
&\quad +\frac{\sum_{i=1}^{n}\exp\left\{-\nu g\left(X_i,\theta\right)\right\}\left\{\left(\mathrm{d}\nu/\mathrm{d}\theta\right)g\left(X_i,\theta\right)+\nu\,\mathrm{d}g\left(X_i,\theta\right)/\mathrm{d}\theta\right\}}{\sum_{i=1}^{n}\exp\left\{-\nu g\left(X_i,\theta\right)\right\}} \\
&= -\frac{\mathrm{d}\nu}{\mathrm{d}\theta}\frac{1}{n}\sum_{i=1}^{n}g\left(X_i,\theta\right)+\nu\frac{\sum_{i=1}^{n}\exp\left\{-\nu g\left(X_i,\theta\right)\right\}\mathrm{d}g\left(X_i,\theta\right)/\mathrm{d}\theta}{\sum_{i=1}^{n}\exp\left\{-\nu g\left(X_i,\theta\right)\right\}} \\
&\quad -\nu n^{-1}\sum_{i=1}^{n}\frac{\mathrm{d}g\left(X_i,\theta\right)}{\mathrm{d}\theta}.
\end{aligned}
$$

Thus, by Eq. (3.1)

$$\left.\frac{\mathrm{d}^2\tilde{l}\left(\theta\right)}{\mathrm{d}\theta^2}\right|_{\theta=\tilde{\theta}}=-\frac{\mathrm{d}\nu}{\mathrm{d}\theta}\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}g\left(X_i,\tilde{\theta}\right)}{\mathrm{d}\theta}=-\frac{\left\{\sum_{i=1}^{n}\mathrm{d}g\left(X_i,\tilde{\theta}\right)/\mathrm{d}\theta\right\}^2}{n\sum_{i=1}^{n}g^2\left(X_i,\tilde{\theta}\right)}<0.$$

Finally, for the Cressie–Read case, $\tilde{l}(\theta) = n^{-1}\sum_{i=1}^{n}\log\hat{w}_i^{\mathrm{CR}}(\theta) = -\{n(\lambda + 1)\}^{-1}\sum_{i=1}^{n}\log\{\mu + \nu g(X_i,\theta)\}$. Then by Eq. (3.3),

$$\left.\frac{\mathrm{d}^2\tilde{l}\left(\theta\right)}{\mathrm{d}\theta^2}\right|_{\theta=\tilde{\theta}}=-\frac{\left\{\sum_{i=1}^{n}\mathrm{d}g\left(X_i,\tilde{\theta}\right)/\mathrm{d}\theta\right\}^2}{n\sum_{i=1}^{n}g^2\left(X_i,\tilde{\theta}\right)}. \qquad\square$$

Let $b = [\{n^{-1}\sum_{i=1}^{n}\mathrm{d}g(X_i,\tilde{\theta})/\mathrm{d}\theta\}^2/\{n^{-1}\sum_{i=1}^{n}g(X_i,\tilde{\theta})^2\}]^{1/2}$. The main result is proved in the next section.

4. Main result

Before stating the main theorem, we need a few notations. We assume that the prior density $\rho(\theta)$ has a Kth-order continuous derivative at $\tilde{\theta}$. Let $\rho_K(\theta) = \rho(\tilde{\theta}) + \rho'(\tilde{\theta})(\theta - \tilde{\theta}) + \cdots + \rho^{(K)}(\tilde{\theta})(\theta - \tilde{\theta})$ the Kth-order Taylor approximation

of the prior density. Further denote the higher-order derivatives of (pseudo) log empirical likelihood as

$$a_{kn}(\theta) = \frac{1}{k!} \sum_{i=1}^{n} \frac{\mathrm{d}^k \tilde{l}(\theta)}{\mathrm{d}\theta^k}, \ k = 3, \ldots, K+3.$$

Define the summation index set

$$I_{i,k} = \left\{ (m_{3,i}, \ldots, m_{K+3,i}) \in \mathbb{N}^K : \sum_{u=3}^{K+3} m_{u,i} = i, \ \sum_{u=3}^{K+3} m_{u,i}(u-2) = k \right\}.$$

Let $y = \sqrt{n}b(\theta - \tilde{\theta})$ be the normalized posterior random variable and

$$
\begin{aligned}
\alpha_k(y,n) &= \frac{1}{k!}\rho^{(k)}\left(\tilde{\theta}\right)\left(\frac{y}{b}\right)^k + \sum_{j=0}^{k-1}\frac{1}{j!}\rho^{(j)}\left(\tilde{\theta}\right) \\
&\quad \times \sum_{i=\lceil (k-j)/(K+1)\rceil}^{k-j} \frac{1}{i!} \sum_{I_{i,k-j}} \binom{i}{m_{3,i}, \ldots, m_{K+3,i}} \\
&\quad \times \prod_{u=3}^{K+3} \left\{ a_{un}\left(\tilde{\theta}\right) \right\}^{m_{u,i}} \left(\frac{y}{b}\right)^{k+2i+j},
\end{aligned}
$$

where, $k = 0, \ldots, K$. For special cases $k = 0, 1$, we have $\alpha_0(y,n) = \rho(\tilde{\theta})$ and $\alpha_1(y,n) = \rho'(\tilde{\theta})y/b + \rho(\tilde{\theta})a_{3n}(\tilde{\theta})(y/b)^3$. Now define $Y_{(1)} = \sqrt{n}b(h_1 - \tilde{\theta})$ and $Y_{(n)} = \sqrt{n}b(h_2 - \tilde{\theta})$ as the normalized lower and upper bounds of the support of the distribution. Now for any $\xi \in (Y_{(1)}, Y_{(n)})$ and $H_n = [h_1, h_2]$, let

$$P_K(\xi,n) = \sum_{k=0}^{K} \left\{ \int_{Y_{(1)}}^{\xi} \alpha_k(y,n) \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y \right\} n^{-k/2}.$$

To control the higher-order error terms, we need the following assumptions.

Assumption 4. For any $(l_1, \ldots, l_j) \subset \{2, \ldots, K+3\}$,

$$E\left\{ \prod_{i=1}^{j} \frac{\mathrm{d}^{l_i} g(X_1, \theta)}{\mathrm{d}\theta^{l_i}} \right\} < \infty.$$

Moreover, we also need an assumption to guarantee the consistency of M-Estimator,

Assumption 5. $g(\cdot, \theta)$ is either bounded or monotone in θ.

Now we state the main theorem of this section.

Theorem 1 (Fundamental Theorem for Expansion). *Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed. Assume the prior density $\rho(\theta)$ has a*

support containing H_n and has $(K+1)$th-order continuous derivative. Under Assumptions 1, 3, 4 and 5, there exist constants $N_1 > 0$ and $M_1 > 0$, such that

$$\left| \int_{Y_{(1)}}^{\xi} \prod_{i=1}^{n} \hat{w}_i \left(\tilde{\theta} + \frac{y}{\sqrt{nb}} \right) \rho \left(\tilde{\theta} + \frac{y}{\sqrt{nb}} \right) \, \mathrm{d}y - P_K(\xi, n) \right| \leq M_1 n^{-(K+1)/2}, \text{ a.s.}$$

(4.1)

for any $n > N_1$ and $\xi \in (Y_{(1)}, Y_{(n)})$.

Proof. See Appendix D. □

This theorem can not only be used to prove asymptotic expansion of the posterior cumulative distribution function, the main result of this paper, but it can also be used to find the asymptotic expansions of the posterior mean, quantiles and many other quantities of interest, as in [12] and [21].

Next we write the posterior cumulative distribution function as

$$\Pi \left(\theta \leq \tilde{\theta} + \frac{\xi}{\sqrt{nb}} \, \middle| \, X_1, \dots, X_n \right) = \frac{\int_{h_1}^{\tilde{\theta} + \xi/\sqrt{nb}} \prod_{i=1}^{n} \hat{w}_i(\theta) \rho(\theta) \, \mathrm{d}\theta}{\int_{h_1}^{h_2} \prod_{i=1}^{n} \hat{w}_i(\theta) \rho(\theta) \, \mathrm{d}\theta}.$$

Moreover, let $R_n = (Y_{(1)}, Y_{(n)})$ and

$$\Phi(\xi \mid R_n) = \frac{\int_{Y_{(1)}}^{\xi} \varphi(y) \, \mathrm{d}y}{\int_{Y_{(1)}}^{Y_{(n)}} \varphi(y) \, \mathrm{d}y},$$

where $\varphi(y)$ is standard normal density, be restricted to R_n. Define polynomial $\gamma_i(\xi, n)$, $i = 1, \dots, n$ recursively as

$$\int_{Y_{(1)}}^{\xi} \alpha_k(y, n) \exp\left(-\frac{y^2}{2} \right) \, \mathrm{d}y = \sum_{j=0}^{k} \left\{ \int_{Y_{(1)}}^{Y_{(n)}} \alpha_j(y, n) \exp\left(-\frac{y^2}{2} \right) \, \mathrm{d}y \right\} \gamma_{k-j}(\xi, n).$$

The first two terms of $\gamma_i(\xi, n)$ are

$$\gamma_0(\xi, n) = \frac{\rho(\tilde{\theta}) \int_{Y_{(1)}}^{\xi} \exp(-y^2/2) \, \mathrm{d}y}{\rho(\tilde{\theta}) \int_{Y_{(1)}}^{Y_{(n)}} \exp(-y^2/2) \, \mathrm{d}y} = \frac{\Phi(\xi) - \Phi(Y_{(1)})}{\Phi(Y_{(n)}) - \Phi(Y_{(1)})} = \Phi(\xi \mid R_n),$$

and

$$\gamma_1(\xi, n) = \frac{\int_{Y_{(1)}}^{\xi} \exp(-y^2/2) \left\{ \rho'(\tilde{\theta}) y/b + \rho(\tilde{\theta}) a_{3n}(\tilde{\theta}) (y/b)^3 \right\} \, \mathrm{d}y}{\rho(\tilde{\theta}) \int_{Y_{(1)}}^{Y_{(n)}} \exp(-y^2/2) \, \mathrm{d}y}$$

$$- \frac{\int_{Y_{(1)}}^{Y_{(n)}} \exp(-y^2/2) \left\{ \rho'(\tilde{\theta}) y/b + \rho(\tilde{\theta}) a_{3n}(\tilde{\theta}) (y/b)^3 \right\} \, \mathrm{d}y}{\rho(\tilde{\theta}) \int_{Y_{(1)}}^{Y_{(n)}} \exp(-y^2/2) \, \mathrm{d}y \, \{ \Phi(\xi \mid R_n) \}^{-1}}$$

$$
= \left\{ \frac{\rho'\left(\tilde{\theta}\right)}{b\rho\left(\tilde{\theta}\right)} \right\} \frac{\varphi\left(Y_{(1)}\right) - \varphi\left(\xi\right) - \Phi\left(\xi \mid R_n\right)\left\{\varphi\left(Y_{(1)}\right) - \varphi\left(Y_{(n)}\right)\right\}}{\int_{Y_{(1)}}^{Y_{(n)}} \varphi\left(y\right) \, dy}
$$

$$
+ \left\{ \frac{a_{3n}\left(\tilde{\theta}\right)}{b^3} \right\} \left\{ \frac{Y_{(1)}^2 \varphi\left(Y_{(1)}\right) + 2\varphi\left(Y_{(1)}\right) - \xi^2 \varphi\left(\xi\right) - 2\varphi\left(\xi\right)}{\int_{Y_{(1)}}^{Y_{(n)}} \varphi\left(y\right) \, dy} \right.
$$

$$
\left. - \Phi\left(\xi \mid R_n\right) \frac{Y_{(1)}^2 \varphi\left(Y_{(1)}\right) + 2\varphi\left(Y_{(1)}\right) - Y_{(n)}^2 \varphi\left(Y_{(n)}\right) - 2\varphi\left(Y_{(n)}\right)}{\int_{Y_{(1)}}^{Y_{(n)}} \varphi\left(y\right) \, dy} \right\}.
$$

We now provide the next important result result of this section, namely the asymptotic expansion of the posterior distribution function.

Theorem 2 (Asymptotic Expansion of the Posterior Cumulative Distribution Function)**.** *Use the same assumptions as in Theorem 1. Then there exist constants N_2 and M_2, such that for any $n \geq N_2$ and $\xi \in (Y_{(1)}, Y_{(n)})$,*

$$
\left| \Pi\left(\theta \leq \tilde{\theta} + \frac{\xi}{\sqrt{n}b} \mid X_1, \ldots, X_n \right) - \Phi\left(\xi \mid R_n\right) - \sum_{i=1}^{K} \gamma_i\left(\xi, n\right) n^{-i/2} \right|
$$
$$
\leq M_2 n^{-(K+1)/2}, \tag{4.2}
$$

almost surely.

Proof. By Theorem 1, we have

$$
\left| \int_{Y_{(1)}}^{\xi} \prod_{i=1}^{n} \hat{w}_i\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) \rho\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) \, dy - P_K\left(\xi, n\right) \right| \leq M_1 n^{-(K+1)/2}, \tag{4.3}
$$

and

$$
\left| \int_{Y_{(1)}}^{Y_{(n)}} \prod_{i=1}^{n} \hat{w}_i\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) \rho\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) \, dy - P_K\left(Y_{(n)}, n\right) \right| \leq M_1 n^{-(K+1)/2}. \tag{4.4}
$$

By definition

$$
\Pi\left(\theta \leq \tilde{\theta} + \frac{\xi}{\sqrt{n}b} \mid X_1, X_2, \ldots, X_n \right)
$$
$$
= \frac{\int_{Y_{(1)}}^{\xi} \prod_{i=1}^{n} \hat{w}_i\left(\tilde{\theta} + y/\sqrt{n}b\right) \rho\left(\tilde{\theta} + y/\sqrt{n}b\right) \, dy}{\int_{Y_{(1)}}^{Y_{(n)}} \prod_{i=1}^{n} \hat{w}_i\left(\tilde{\theta} + y/\sqrt{n}b\right) \rho\left(\tilde{\theta} + y/\sqrt{n}b\right) \, dy}.
$$

We know that all the terms in Eq. (4.3) and Eq. (4.4), are integrals of continuous functions over bounded closed intervals. Hence they are almost surely bounded below by some constant C_1 and bounded above by some constant C_2, for all $n > N_1$. Then

$$
\left| \Pi\left(\theta \leq \tilde{\theta} + \frac{\xi}{\sqrt{n}b} \mid X_1, X_2, \ldots, X_n \right) - \frac{P_K\left(\xi, n\right)}{P_K\left(Y_{(n)}, n\right)} \right|
$$

$$
\begin{aligned}
&= \left| \frac{\int_{Y_{(1)}}^{\xi} \prod_{i=1}^{n} \hat{w}_i \left(\tilde{\theta} + y/\sqrt{n}b \right) \rho \left(\tilde{\theta} + y/\sqrt{n}b \right) \mathrm{d}y}{\int_{Y_{(1)}}^{Y_{(n)}} \prod_{i=1}^{n} \hat{w}_i \left(\tilde{\theta} + y/\sqrt{n}b \right) \rho \left(\tilde{\theta} + y/\sqrt{n}b \right) \mathrm{d}y} - \frac{P_K \left(\xi, n \right)}{P_K \left(Y_{(n)}, n \right)} \right| \\
&= \left| \frac{\int_{Y_{(1)}}^{\xi} \prod_{i=1}^{n} \hat{w}_i \left(\tilde{\theta} + y/\sqrt{n}b \right) \rho \left(\tilde{\theta} + y/\sqrt{n}b \right) \mathrm{d}y - P_K \left(\xi, n \right)}{\int_{Y_{(1)}}^{Y_{(n)}} \prod_{i=1}^{n} \hat{w}_i \left(\tilde{\theta} + y/\sqrt{n}b \right) \rho \left(\tilde{\theta} + y/\sqrt{n}b \right) \mathrm{d}y} \right. \\
&\quad \left. - \frac{\left\{ \int_{Y_{(1)}}^{Y_{(n)}} \prod_{i=1}^{n} \hat{w}_i \left(\tilde{\theta} + y/\sqrt{n}b \right) \rho \left(\tilde{\theta} + y/\sqrt{n}b \right) \mathrm{d}y - P_K \left(Y_{(n)}, n \right) \right\} P_K \left(\xi, n \right)}{\left\{ \int_{Y_{(1)}}^{Y_{(n)}} \prod_{i=1}^{n} \hat{w}_i \left(\tilde{\theta} + y/\sqrt{n}b \right) \rho \left(\tilde{\theta} + y/\sqrt{n}b \right) \mathrm{d}y \right\} P_K \left(Y_{(n)}, n \right)} \right| \\
&\leq \frac{1}{\left| \int_{Y_{(1)}}^{Y_{(n)}} \prod_{i=1}^{n} \hat{w}_i \left(\tilde{\theta} + y/\sqrt{n}b \right) \rho \left(\tilde{\theta} + y/\sqrt{n}b \right) \mathrm{d}y \right|} \tag{4.5} \\
&\quad \times \left\{ \left| \int_{Y_{(1)}}^{\xi} \prod_{i=1}^{n} \hat{w}_i \left(\tilde{\theta} + \frac{y}{\sqrt{n}b} \right) \rho \left(\tilde{\theta} + \frac{y}{\sqrt{n}b} \right) \mathrm{d}y - P_K \left(\xi, n \right) \right| \right. \tag{4.6} \\
&\quad \left. + \left| \int_{Y_{(1)}}^{Y_{(n)}} \prod_{i=1}^{n} \hat{w}_i \left(\tilde{\theta} + \frac{y}{\sqrt{n}b} \right) \rho \left(\tilde{\theta} + \frac{y}{\sqrt{n}b} \right) \mathrm{d}y - P_K \left(Y_{(n)}, n \right) \right| \left| \frac{P_K \left(\xi, n \right)}{P_K \left(Y_{(n)}, n \right)} \right| \right\} \\
&\leq \frac{1}{C_1} \left\{ M_1 n^{-(K+1)/2} + M_1 n^{-(K+1)/2} \frac{C_2}{C_1} \right\} = \frac{M_1}{C_1} \left(1 + \frac{C_2}{C_1} \right) n^{-(K+1)/2}. \tag{4.7}
\end{aligned}
$$

Now we find the quotient series of $P_K(\xi, n)/P_K(Y_{(n)}, n)$. By the definition of $\gamma_i(\xi, n)$, through simple calculation, we have

$$
\frac{P_K \left(\xi, n \right)}{P_K \left(Y_{(n)}, n \right)} = \sum_{i=0}^{\infty} \gamma_i \left(A, n \right) n^{-i/2},
$$

By the discussion following Lemma 7 in the Appendix B , we know that all γ_i are almost surely uniformly bounded for all large n. Thus, there exists a constant M_3, such that

$$
\left| \frac{P_K \left(\xi, n \right)}{P_K \left(Y_{(n)}, n \right)} - \Phi \left(\xi \mid R_n \right) - \sum_{i=1}^{K} \gamma_i \left(A, n \right) n^{-i/2} \right| \leq M_3 n^{-(K+1)/2}. \tag{4.8}
$$

We combine Eq. (4.7) and Eq. (4.8), to get Eq. (4.2). $\qquad \square$

Let $K = 2$. Then we get asymptotic normality of the posterior distribution.

Corollary 1 (Bernstein-von Mises Theorem)**.** *Use the assumption in Theorem 1 with $K = 2$, then the posterior distribution converges in distribution to normal distribution almost surely, that is*

$$
\sqrt{n}b \left(\theta - \tilde{\theta} \right) \Big| X_1, \ldots, X_n \to N(0, 1), \text{ a.s.},
$$

Theorem 1 builds a strong foundation for asymptotic expansions of many other quantities based on the posterior, such as the mean and higher-order

posterior moments. This follows simply by replacing the prior density $\rho(\theta)$ with an appropriate function. Here we use the posterior mean as an example. More examples can be found in [12].

Example 1. Replace $\rho(\theta)$ in Eq. (4.1) by $y\rho(\theta)$,

$$\left| \int_{Y_{(1)}}^{Y_{(n)}} \prod_{i=1}^{n} \hat{w}_i \left(\tilde{\theta} + \frac{y}{\sqrt{nb}} \right) \left\{ y\rho \left(\tilde{\theta} + \frac{y}{\sqrt{nb}} \right) \right\} \, \mathrm{d}y - P_K^N \left(Y_{(n)}, n \right) \right| \leq M_1 n^{-(K+1)/2},$$

where

$$P_K^N (\xi, n) = \sum_{k=0}^{K} \left\{ \int_{Y_{(1)}}^{\xi} \alpha_k (y, n) \exp \left(-\frac{y^2}{2} \right) y \, \mathrm{d}y \right\} n^{-(K+1)/2}.$$

Applying the same argument as in the proof of Theorem 2, the asymptotic expansion of the posterior mean is

$$E \left\{ \sqrt{nb} \left(\theta - \tilde{\theta} \right) \mid X \right\} = \left\{ \frac{\rho' \left(\tilde{\theta} \right)}{\rho \left(\tilde{\theta} \right) b} \frac{\int_{Y_{(1)}}^{Y_{(n)}} y^2 \varphi (y) \, \mathrm{d}y}{\int_{Y_{(1)}}^{Y_{(n)}} \varphi (y) \, \mathrm{d}y} + \frac{a_{3n}}{b^3} \frac{\int_{Y_{(1)}}^{Y_{(n)}} y^4 \varphi (y) \, \mathrm{d}y}{\int_{Y_{(1)}}^{Y_{(n)}} \varphi (y) \, \mathrm{d}y} \right\} n^{-1}$$
$$+ O_P \left(n^{-\frac{3}{2}} \right). \qquad (4.9)$$

Since $Y_{(n)} \to +\infty$ and $Y_{(1)} \to -\infty$ a.s. as $n \to \infty$,

$$\lim_{n \to \infty} \frac{\int_{Y_{(1)}}^{Y_{(n)}} y^2 \varphi (y) \, \mathrm{d}y}{\int_{Y_{(1)}}^{Y_{(n)}} \varphi (y) \, \mathrm{d}y} = \int_{\mathbb{R}} y^2 \varphi (y) \, \mathrm{d}y = 1,$$

and

$$\lim_{n \to \infty} \frac{\int_{Y_{(1)}}^{Y_{(n)}} y^4 \varphi (y) \, \mathrm{d}y}{\int_{Y_{(1)}}^{Y_{(n)}} \varphi (y) \, \mathrm{d}y} = \int_{\mathbb{R}} y^4 \varphi (y) \, \mathrm{d}y = 3.$$

Then a moment matching prior ([9]) is found as the solution of

$$\frac{\rho' (\theta)}{\rho (\theta)} = - \lim_{n \to \infty} \frac{3a_{3n}}{b^2}.$$

For the empirical likelihood and the exponentially tilted empirical likelihood, some heavy algebra yields

$$a_{3n} \left(\tilde{\theta} \right) = \frac{\left\{ \sum_{i=1}^{n} \partial g \left(X_i, \tilde{\theta} \right) / \partial \theta \right\}^3 \sum_{i=1}^{n} g^3 \left(X_i, \tilde{\theta} \right)}{3n \left\{ \sum_{i=1}^{n} g^2 \left(X_i, \tilde{\theta} \right) \right\}^3}$$
$$- \frac{\left\{ \sum_{i=1}^{n} \partial g \left(X_i, \tilde{\theta} \right) / \partial \theta \right\}^2 \sum_{i=1}^{n} g \left(X_i, \tilde{\theta} \right) \partial g \left(X_i, \tilde{\theta} \right) / \partial \theta}{n \left\{ \sum_{i=1}^{n} g^2 \left(X_i, \tilde{\theta} \right) \right\}^2}$$

$$+\frac{\sum_{i=1}^{n}\partial g\left(X_i,\tilde{\theta}\right)/\partial\theta\sum_{i=1}^{n}\partial^2 g\left(X_i,\tilde{\theta}\right)/\partial\theta^2}{2n\left\{\sum_{i=1}^{n}g^2\left(X_i,\tilde{\theta}\right)\right\}}.$$

Using strong law of large numbers,

$$\lim_{n\to\infty}a_{3n}\;=\;\frac{\left[E\left\{\partial g\left(X_1,\theta\right)/\partial\theta\right\}\right]^3 E\left\{g^3\left(X_1,\theta\right)\right\}}{3\left[E\left\{g^2\left(X_1,\theta\right)\right\}\right]^3}$$

$$-\frac{\left[E\left\{\partial g\left(X_1,\theta\right)/\partial\theta\right\}\right]^2 E\left\{g\left(X_1,\theta\right)\partial g\left(X_1,\theta\right)/\partial\theta\right\}}{\left\{Eg^2\left(X_1,\theta\right)\right\}^2}$$

$$+\frac{E\left\{\partial g\left(X_1,\theta\right)/\partial\theta\right\}E\left\{\partial^2 g\left(X_1,\theta\right)/\partial\theta^2\right\}}{2E\left\{g^2\left(X_1,\theta\right)\right\}}\;\text{a.s. }(P_\theta).$$

Hence, we have the following corollary.

Corollary 2. *Assume the conditions in Theorem 1 are satisfied at $K = 4$. Then the first order moment matching prior of Bayesian empirical likelihood and Bayesian exponentially tilted empirical likelihood is*

$$\rho\left(\theta\right)\;=\;\exp\Bigg\{-\int_{-\infty}^{\theta}\Bigg(\frac{\left[E\left\{\partial g\left(X_1,s\right)/\partial s\right\}\right]^3 E\left\{g^3\left(X_1,s\right)\right\}}{\left[E\left\{g^2\left(X_1,s\right)\right\}\right]^4}$$

$$+\frac{3\left[E\left\{\partial g\left(X_1,s\right)/\partial s\right\}\right]^2 E\left\{g\left(X_1,s\right)\partial g\left(X_1,s\right)/\partial s\right\}}{\left[E\left\{g^2\left(X_1,s\right)\right\}\right]^3}$$

$$-\frac{3E\left\{\partial g\left(X_1,s\right)/\partial s\right\}E\left\{\partial^2 g\left(X_1,s\right)/\partial s^2\right\}}{2\left[E\left\{g^2\left(X_1,s\right)\right\}\right]^2}\Bigg)\,\mathrm{d}s\Bigg\}.$$

In the special case, $g(x,\theta) = x - \theta$, by Corollary 2, the moment matching prior is

$$\rho\left(\theta\right)=\exp\left(\int_{-\infty}^{\theta}\frac{E\left\{(X_1-s)^3\right\}}{\left[E\left\{(X_1-s)^2\right\}\right]^4}\,\mathrm{d}s\right).$$

Remark 1. Because the posterior mean is the most widely used Bayesian estimator, this result could provide a useful tools for higher order analysis. For example, our result would produce the same result in [20], which estimates the quantile using the posterior mean of Bayesian empirical likelihood based on smoothed estimating equation.

Let $K(x) = \int_{-\infty}^{x}k(u)\,\mathrm{d}u$ be a kernel smoothing function, and $K_h(\cdot) = K(\cdot/h)h$. The smoothed constraint of the empirical likelihoods to estimate the α quantile θ is

$$\sum_{i=1}^{n}w_i\left[K_h\left(\theta-X_i\right)-(1-\alpha)\right]=0,$$

and $g(\theta, X_i) = K_h(\theta - X_i) - (1 - \alpha)$. Hence the asymptotic variance is

$$
\begin{aligned}
b^2 &= \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial g(\theta, X_i)}{\partial \theta} \Big|_{\theta = \tilde{\theta}} \right) \Big/ \frac{1}{n} \sum_{i=1}^{n} g^2\left(\tilde{\theta}, X_i\right) \\
&= \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial K_h(\theta - X_i)}{\partial \theta} \Big|_{\theta = \tilde{\theta}} \right) \Big/ \frac{1}{n} \sum_{i=1}^{n} \left[K_h\left(\tilde{\theta} - X_i\right) - (1 - \alpha) \right]^2 \\
&= \left(\frac{1}{n} \sum_{i=1}^{n} k_h\left(\tilde{\theta} - X_i\right) \right) \Big/ \frac{1}{n} \sum_{i=1}^{n} \left[K_h\left(\tilde{\theta} - X_i\right) - (1 - \alpha) \right]^2.
\end{aligned}
$$

If the prior on θ is normal with parameters μ_0 and σ_0, then

$$
\frac{\rho'(\theta)}{\rho(\theta)} \Big|_{\theta = \tilde{\theta}} = -\frac{\tilde{\theta} - \mu_0}{\sigma_0^2}.
$$

Hence, by Eq. 4.9, we have

$$
E\left(\theta \mid X_1, \ldots, X_n\right) \approx \tilde{\theta} - \frac{\tilde{\theta} - \mu_0}{\sigma_0^2 b^2} n^{-1},
$$

which is the same as the formula in Remark 1 of [20].

5. Simulation results

In this section, we give some simulation results.

5.1. Heavy tailed distribution

First we take $g(X_i, \theta) = X_i - \theta$, $i = 1, \ldots, n$. Let $K = 3$, we compare the first order approximation with normal approximation and second order approximation. By heavy algebra, we get for all the three empirical likelihoods,

$$
\tilde{l}^{(3)}\left(\overline{X}\right) = \frac{2n^2 \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^3}{\left\{ \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \right\}^3}.
$$

So $\tilde{l}^{(3)}(\overline{X})$ for the three empirical likelihoods are asymptotically equivalent up to the second order. The true cumulative distribution function is calculated by numerical integration. The normal approximation polynomial is $\Phi(\xi \mid R_n)$, and the second order approximation polynomial is

$$
\begin{aligned}
\Phi\left(\xi \mid R_n\right) &+ \frac{1}{\sqrt{n}} \left[\left\{ \frac{\rho'\left(\overline{X}\right)}{\rho\left(\overline{X}\right) b} \right\} \right. \\
&\times \left\{ \frac{\varphi\left(Y_{(1)}\right) - \varphi(\xi)}{\int_{Y_{(1)}}^{Y_{(n)}} \varphi(y)\, dy} - \frac{\varphi\left(Y_{(1)}\right) - \varphi\left(Y_{(n)}\right)}{\int_{Y_{(1)}}^{Y_{(n)}} \varphi(y)\, dy} \Phi\left(\xi \mid R_n\right) \right\}
\end{aligned}
$$

$$+ \left\{ \frac{2n^{-1} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^3}{6b^9} \right\} \left\{ \frac{Y_{(1)}^2 \varphi \left(Y_{(1)} \right) + 2\varphi \left(Y_{(1)} \right) - \xi^2 \varphi \left(\xi \right) - 2\varphi \left(\xi \right)}{\int_{Y_{(1)}}^{Y_{(n)}} \varphi \left(y \right) \, \mathrm{d}y} \right.$$

$$\left. - \frac{Y_{(1)}^2 \varphi \left(Y_{(1)} \right) + 2\varphi \left(Y_{(1)} \right) - Y_{(n)}^2 \varphi \left(Y_{(n)} \right) - 2\varphi \left(Y_{(n)} \right)}{\int_{Y_{(1)}}^{Y_{(n)}} \varphi \left(y \right) \, \mathrm{d}y} \Phi \left(\xi \mid R_n \right) \right\} \right].$$

We take samples of size $n = 10$, and 80 from a t distribution with degrees of freedom 6, and the Cauchy prior. Set Cressie–Read divergence parameter $\lambda = 2$. Then

$$\rho \left(\overline{X} \right) = \frac{1}{\pi \left(1 + \overline{X}^2 \right)},$$

$$\rho' \left(\overline{X} \right) = -\frac{2\overline{X}}{\pi \left(1 + \overline{X}^2 \right)^2}.$$

The results are given in Figure 1 on page 3027 and Figure 2 on page 3028. In both plots, the red line stands for normal approximation of the posterior cumulative distribution function, the blue line stands for the first order approximation, the green line stands for the posterior based on the empirical likelihood, the purple line stands for the posterior based on the exponentially tilted empirical likelihood, and the black line stands for the Cressie–Read divergence empirical likelihood. We see that even when the sample size is 10, the three types of empirical likelihoods are quite close to each other, which supports the fact they are equivalent at least up to the second order, and the second order approximation works well. The first order approximation is closer than the normal approximation, which lends credence to our theorem. When the sample size increases to 80, all the lines almost coincide with each other, which means that the approximations are quite successful.

5.2. Skewed distribution and different priors

Here we use gamma distribution with shape parameter 2 and scale parameter 0.2, so that the skewness is $2/\sqrt{2} = \sqrt{2}$ and the mean is 0.4. We stil use one dimension constraint $g(X, \theta) = X - \theta$ to estimate mean. The priors are Cauchy distributions with different means μ_0 and different variances σ_0^2. We use the accuracy defined as

$$E_X \left(\max_y \left| P \left(\theta \leq y \mid X \right) - \tilde{P} \left(\theta \leq y \mid X \right) \right| \right)$$

to measure the performance of our approximations $\tilde{P}(\theta \leq y \mid X)$ with respect to true Bayesian empirical likelihood posteriors. The results are summarized in Table 1. The table is organized as follows:

- The two columns under Empirical Likelihood means the true posteriors are based on the empirical likelihood.
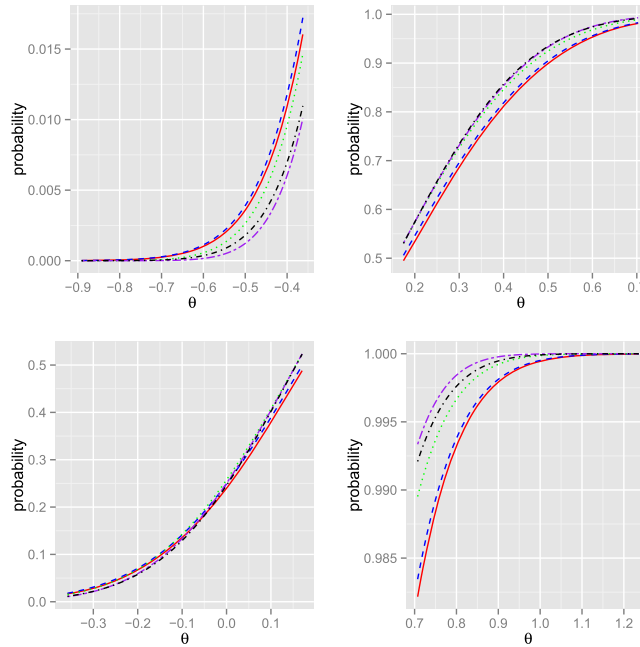
F$_{\text{IG}}$ 1. *Posterior cumulative distribution function when sample size is 10*

– The Normal column under Empirical Likelihood column documents the simulated accuracies when we use normal approximations to estimate the true posteriors based on the empirical likelihoods.

– The 1st Order column under Empirical Likelihood documentclass the simulated accuracies when we use first order approximations to estimate the true posteriors based on the empirical likelihoods.

Other columns can be intepretted similarly. The table justifies our theoretical results well.

5.3. Comparison with parametric Bayesian model

By our Corollary 1, we know the asymptotic variance of the Bayesian empirical likelihood is the inverse of Godambe information number, while in parametric Bayesian model, the asymptotic variance is the inverse of Fisher information number. Generally speaking, the Fisher information number will be larger than the Godambe information number. The difference between the asymptotic variances serves as a "payment" to use a semiparametric model instead of a full parametric model. We also do some simulation to further justify this statements. Let $X_i \sim \exp(\lambda_0)$, be i.i.d samples, where $EX_1 = \lambda_0^{-1} = 1$. Let the prior on λ be a gamma distribution with shape parameter $\alpha = 2$ and scale parameter $\beta = 1$. The the parametric posterior of λ is $\Gamma(\alpha + n, (\beta^{-1} + \sum_{i=1}^{n} X_i)^{-1})$. The
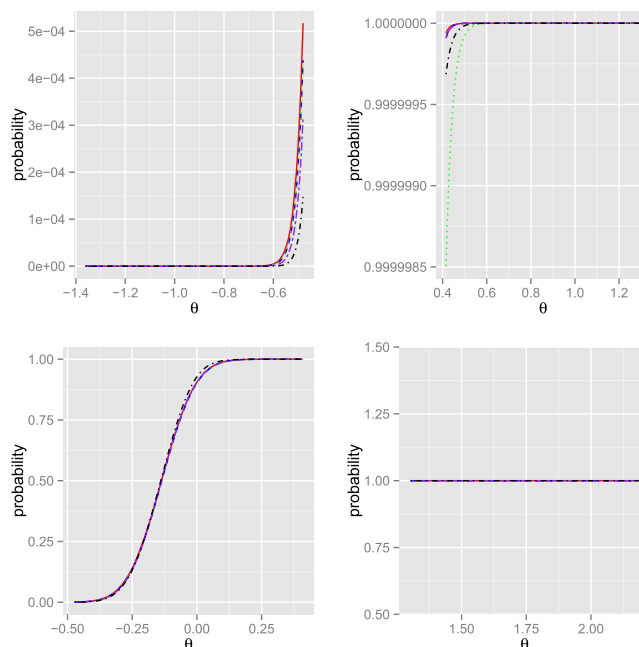
FIG 2. *Posterior cumulative distribution function when sample size is 80*

parameter of interest is still the population mean, which is inverse of λ_0. We compare the full parametric approaches under the true model and the misspecified model and the empirical likelihood approaches by measuring the Bayesian risk of the mean under square loss. In the full parametric approach under the true model, we can directly calculate the theoretical Bayes risk,

$$E_X E_\lambda \left(\left(\lambda^{-1} - \lambda_0^{-1} \right)^2 \mid X \right) = \frac{(1+n)^2 + n}{n(n+1)^2} + \frac{n}{(n+1)^2}.$$

In the full parametric approach under the misspecified model, we assume the data are drawn from χ_ν^2. We use inverse gamma distribution with $\alpha = 2$ and $\beta = 1$ as the prior on ν, so that prior distribution on the population mean are the same under the two full parametric approaches and the three empirical likelihood approaches. The posterior distribution under this prior is proportional to

$$\left[\Gamma \left(\frac{\nu}{2} \right) \right]^{-n} \nu^{-3} \exp \left(-\frac{1}{\nu} - \frac{n \log 2 - \sum_{i=1}^n \log X_i}{2} \nu \right)$$

$$= \left[\Gamma \left(\frac{\nu}{2} \right) \right]^{-n} \left(\prod_{i=1}^n X_i \right)^{\nu/2} \nu^{-3/2} \exp \left(-\sqrt{2n \log 2} \right)$$

TABLE 1
Accuracy of Approximation under Different Priors and Different Sample Sizes

$n =$	$\mu_0 =$	$\sigma_0 =$	Empirical Likelihood		Exponentially Tilted		Cressie–Read	
			Normal	1st Order	Normal	1st Order	Normal	1st Order
		0.3	0.041665	0.056081	0.061981	0.071276	0.045879	0.061546
	-1	1	0.042696	0.055215	0.062289	0.070304	0.047265	0.060279
		3	0.049532	0.053994	0.065669	0.068312	0.053553	0.058001
		0.3	0.133635	0.069615	0.115206	0.070635	0.132666	0.068341
10	1	1	0.073574	0.054025	0.077864	0.065787	0.075428	0.056252
		3	0.056341	0.053884	0.068747	0.067331	0.059624	0.057174
		0.3	0.058219	0.054058	0.069527	0.067091	0.061305	0.057170
	10	1	0.058175	0.054056	0.069506	0.067096	0.061265	0.057171
		3	0.057821	0.054044	0.069343	0.067135	0.060941	0.057185
		0.3	0.033282	0.049899	0.036867	0.052868	0.035582	0.052853
	-1	1	0.036950	0.050032	0.039354	0.051875	0.039047	0.052437
		3	0.046609	0.050783	0.046149	0.050188	0.047747	0.051989
		0.3	0.119071	0.067804	0.107920	0.053847	0.116684	0.063350
20	1	1	0.070491	0.053560	0.065260	0.048196	0.069826	0.052368
		3	0.053840	0.051591	0.051581	0.049409	0.054196	0.051921
		0.3	0.055715	0.051957	0.052896	0.049275	0.055817	0.052014
	10	1	0.055670	0.051951	0.052863	0.049279	0.055777	0.052013
		3	0.055314	0.051902	0.052597	0.049307	0.055457	0.052005
		0.3	0.024996	0.036882	0.023464	0.035726	0.023722	0.035670
	-1	1	0.028513	0.037304	0.026368	0.035542	0.026642	0.035604
		3	0.035794	0.038411	0.032585	0.035371	0.032902	0.035626
		0.3	0.080418	0.048038	0.076356	0.041320	0.076594	0.041857
50	1	1	0.050983	0.040794	0.047296	0.035954	0.047471	0.036357
		3	0.040559	0.039186	0.036905	0.035419	0.037221	0.035770
		0.3	0.041696	0.039442	0.037936	0.035484	0.038245	0.035850
	10	1	0.041668	0.039437	0.037910	0.035482	0.038219	0.035848
		3	0.041450	0.039398	0.037703	0.035473	0.038014	0.035837

$$\times \left\{ \nu^{-3/2} \exp\left(-\frac{2\left[\nu - \sqrt{2/\left(n\log 2\right)}\right]^2}{2\left[2/\left(n\log 2\right)\right]\nu} \right) \right\}.$$

The last line of above equations are the kernel of inverse Gaussion distribution with mean parameter $\mu = \sqrt{2/(n\log 2)}$ and scale parameter $\lambda = 2$. Let V be the inverse Gaussion random variable with above parameters, we can calculate the Bayes risk using Monte Calro integration by

$$E_{X_1,\ldots,X_n} \left(\frac{E_V\left\{ (V-1)^2 \left[\Gamma\left(V/2\right)\right]^n \left(\prod_{i=1}^n X_i\right)^{V/2} V^{-3/2} \right\}}{E_V\left\{ \left[\Gamma\left(V/2\right)\right]^n \left(\prod_{i=1}^n X_i\right)^{V/2} V^{-3/2} \right\}} \right).$$

In the empirical likelihood approaches, we need the simulation. The results are summarized in Table 2. We can see the Bayes risks of all the tree empirical likelihoods are larger than the full parametric approach under the true model, but are smaller than that of misspecified model, which substantiates the robustness of our nonparametric approaches against model misspecification.

	EL	ETEL	CREL	True	Misspecified
n=10	0.1602696	0.1560124	0.1590662	0.0586777	0.7499204
n=20	0.0906427	0.0883737	0.0895608	0.0159864	0.7860900
n=50	0.0366288	0.0357060	0.0360126	0.0026990	0.8293515
n=100	0.0192969	0.0189246	0.0189719	0.0006872	0.8571012

6. Discussion

The paper provides an asymptotic expansion of the posterior based on an empirical likelihood subject to a linear constraint. The Bernstein –von Mises theorem and asymptotic expansions of the cumulative distribution function and the posterior mean are obtained as corollaries. Future work will include an extension to the multivariate case as well as expansions subject to multiple constraints. Another potential topic of research is asymptotic expansion of posteriors under regression constraints, extending the arguments of [7, 6]

Appendix A: Behavior of log empirical likelihood in the tail

The Taylor expansion consists of expanding the log empirical likelihood and prior density around the mean and then control the tail part of the log empirical likelihood. In order to implement this idea, the tail part of $\tilde{l}(\theta)$ must vanish faster than the required polynomial order. In this section, we will show that indeed the tail part of $\tilde{l}(\theta)$ vanishes at an exponential rate.

Lemma 5. *Under Assumptions 1 and 2, for any $\delta_1 > 0$, there exist $\varepsilon_1 > 0$ and N_3, such that*

$$\tilde{l}(\theta) - \tilde{l}\left(\tilde{\theta}\right) \leq -\varepsilon, \text{ a.s.},$$

for any $|b(\theta - \tilde{\theta})| \geq \delta_1$ and $\theta \in H_n$, where $b = [\{n^{-1}\sum_{i=1}^{n} \mathrm{d}g(X_i, \tilde{\theta})/\mathrm{d}\theta\}^2/\{n^{-1}\sum_{i=1}^{n} g(X_i, \tilde{\theta})^2\}]^{-1/2}$

Proof. By Lemma 3, we know that $\tilde{\theta}$ is the unique maximizer. Therefore, for any $\theta \neq \tilde{\theta}$, $\tilde{l}(\theta) < \tilde{l}(\tilde{\theta})$. The set

$$\left\{\theta : \left|b\left(\theta - \tilde{\theta}\right)\right| \geq \delta_1\right\} \cap H_n$$

is a compact set, and $\tilde{l}(\theta)$ is a continuous function. Hence there exists a $\theta^* \in \{\theta : |b(\theta - \tilde{\theta})| \geq \delta_1\} \cap H_n$, such that for any $\theta \in \{\theta : |b(\theta - \tilde{\theta})| \geq \delta_1\} \cap H_n$,

$$\tilde{l}(\theta) \leq \tilde{l}(\theta^*).$$

Therefore, $\tilde{l}(\theta) \leq \tilde{l}(\theta^*) < \tilde{l}(\tilde{\theta})$, which is equivalent to

$$\tilde{l}(\theta) - \tilde{l}\left(\tilde{\theta}\right) \leq \tilde{l}(\theta^*) - \tilde{l}\left(\tilde{\theta}\right) < 0.$$

Let $\varepsilon_1 = \{\tilde{l}(\tilde{\theta}) - \tilde{l}(\theta^*)\}/2$, then we have

$$\tilde{l}(\theta) - \tilde{l}\left(\tilde{\theta}\right) \leq \tilde{l}(\theta^*) - \tilde{l}\left(\tilde{\theta}\right) < \varepsilon_1. \qquad \square$$

Appendix B: Higher-order derivatives

In order to expand around the mean, we need to control the remainder terms in the Taylor expansion, which involves the finiteness of higher-order derivatives of $\tilde{l}$.

Lemma 6. *Let*

$$\omega_i(\theta) = \begin{cases} \{1 + \nu(\theta)g(X_i, \theta)\}^{-1}, & \text{for EL,} \\ \exp\{-\nu(\theta)g(X_i, \theta)\}, & \text{for ETEL,} \\ \{\mu(\theta) + \nu(\theta)g(X_i, \theta)\}^{-1/(\lambda+1)}, & \text{for CREL.} \end{cases}$$

and

$$D = \begin{cases} n^{-1}\sum_{i=1}^n \omega_i^r g(X_i, \theta)^2, & \text{for EL and ETEL,} \\ n^{-1}\sum_{i=1}^n \omega_i^{\lambda+2}\left[g(X_i, \theta) - \{n^{-1}\sum_{i=1}^n \omega_i^{\lambda+2}g(X_i, \theta)\}\right]^2, & \text{for CREL.} \end{cases}$$

Then under Assumptions 1 and 3, for any $k = 2, \ldots K+3$,

$$\frac{\mathrm{d}^k}{\mathrm{d}\theta^k}\tilde{l}(\theta) = D^{-r_k}P_k(M_1, M_2, \ldots, \nu, \mu),$$

where P_k are polynomial function, and all the $r_j < C(k)$, and $C(k)$ is some constant only depending on k, and M_j are the weighted average of higher order derivatives of g, i.e.

$$M_j = \frac{1}{n}\sum_{i=1}^n \omega_i^r \prod_l \frac{\mathrm{d}^l g(X_i, \theta)}{\mathrm{d}\theta^l},$$

l can be the same.

Proof. From Lemma 2, we know that

$$\frac{\mathrm{d}\nu}{\mathrm{d}\theta} = D^{-1}P_1\left\{\frac{1}{n}\sum_{i=1}^n \omega_i^{r_1}\frac{\mathrm{d}g(X_i, \theta)}{\mathrm{d}\theta}, \frac{1}{n}\sum_{i=1}^n \omega_i^{r_2}g(X_i, \theta), \ldots, \nu, \mu\right\}.$$

Moreover,

$$\begin{aligned} \frac{\mathrm{d}\tilde{l}^{\mathrm{EL}}(\theta)}{\mathrm{d}\theta} &= \sum_{i=1}^n \frac{1}{1+\nu g(X_i, \theta)}\left\{\frac{\mathrm{d}\nu}{\mathrm{d}\theta}g(X_i, \theta) + \nu\frac{\mathrm{d}g(X_i, \theta)}{\mathrm{d}\theta}\right\} \\ &= \frac{\mathrm{d}\nu}{\mathrm{d}\theta}\frac{1}{n}\sum_{i=1}^n \omega_i(\theta)g(X_i, \theta) + \nu\frac{1}{n}\sum_{i=1}^n \omega_i(\theta)\frac{\mathrm{d}g(X_i, \theta)}{\mathrm{d}\theta}, \end{aligned}$$

$$\frac{\mathrm{d}\tilde{l}^{\mathrm{ET}}(\theta)}{\mathrm{d}\theta} = -\frac{\mathrm{d}\nu}{\mathrm{d}\theta}\frac{1}{n}\sum_{i=1}^{n}g(X_i,\theta) - \nu\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}g(X_i,\theta)}{\mathrm{d}\theta} + \nu\sum_{i=1}^{n}\hat{w}_i(\theta)\frac{\mathrm{d}g(X_i,\theta)}{\mathrm{d}\theta},$$

$$\frac{\mathrm{d}\tilde{l}^{\mathrm{CR}}(\theta)}{\mathrm{d}\theta} = -\frac{1}{\lambda+1}\sum_{i=1}^{n}\omega_i^{\lambda+1}\left\{\frac{\mathrm{d}\mu}{\mathrm{d}\theta} + \frac{\mathrm{d}\nu}{\mathrm{d}\theta}g(X_i,\theta) + \nu\frac{\mathrm{d}g(X_i,\theta)}{\mathrm{d}\theta}\right\}$$

$$= -\frac{1}{\lambda+1}\left\{\frac{\mathrm{d}\mu}{\mathrm{d}\theta}\frac{1}{n}\sum_{i=1}^{n}\omega_i^{\lambda+1} + \frac{\mathrm{d}\nu}{\mathrm{d}\theta}\frac{1}{n}\sum_{i=1}^{n}\omega_i^{\lambda+1}g(X_i,\theta)\right.$$

$$\left. + \nu\frac{1}{n}\sum_{i=1}^{n}\omega_i^{\lambda+1}\frac{\mathrm{d}g(X_i,\theta)}{\mathrm{d}\theta}\right\}.$$

So for $k = 1$, the lemma holds. Assume for $k = n$, the lemma holds. Then for $k = n + 1$,

$$\frac{\mathrm{d}^{n+1}\tilde{l}(\theta)}{\mathrm{d}\theta^{n+1}} = \frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\mathrm{d}^k\tilde{l}(\theta)}{\mathrm{d}\theta}$$

$$= -r_k D^{-r_k-1}\left(\frac{\mathrm{d}D}{\mathrm{d}\theta}\sum_{i=1}^{k_n}\frac{\mathrm{d}P_k}{\mathrm{d}M_i}\frac{\mathrm{d}M_i}{\mathrm{d}\theta} + \frac{\mathrm{d}P_k}{\mathrm{d}\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\theta} + \frac{\mathrm{d}P_k}{\mathrm{d}\mu}\frac{\mathrm{d}\mu}{\mathrm{d}\theta}\right)$$

The partial derivative of P_k is still a polynomial. Also, D itself is a polynomial in $n^{-1}\sum_{i=1}^{n}\omega_i^r g(X_i,\theta)^2$, μ, and ν. Next

$$\frac{\mathrm{d}M_i}{\mathrm{d}\theta} = \frac{1}{n}\sum_{i=1}^{n}\left\{r\omega_i^{r-1}\frac{\mathrm{d}\omega_i}{\mathrm{d}\theta}\prod_l\frac{\mathrm{d}^l g(X_i,\theta)}{\mathrm{d}\theta^l}\right.$$

$$\left. + \omega_i^r\sum_{l_j}\prod_{l\neq l_j}\frac{\mathrm{d}^l g(X_i,\theta)}{\mathrm{d}\theta^l}\frac{\mathrm{d}^{l_j+1}g(X_i,\theta)}{\mathrm{d}\theta^{l_j+1}}\right\}$$

$$= \frac{r}{n}\sum_{i=1}^{n}\omega_i^{r-1}\prod_l\frac{\mathrm{d}^l g(X_i,\theta)}{\mathrm{d}\theta^l}\frac{\mathrm{d}\omega_i}{\mathrm{d}\theta}$$

$$+ \sum_{l_j}\frac{1}{n}\sum_{i=1}^{n}\omega_i^r\prod_{l\neq l_j}\frac{\mathrm{d}^l g(X_i,\theta)}{\mathrm{d}\theta^l}\frac{\mathrm{d}^{l_j+1}g(X_i,\theta)}{\mathrm{d}\theta^{l_j+1}}.$$

Similar to the calculation of the first order derivative of empirical log likelihood, we know the $\mathrm{d}\omega_i/\mathrm{d}\theta$ are polynomials involving terms like M_i. So for $k = n+1$, the higher order derivatives of empirical log likelihood are of a similar form. Hence, by mathematical induction, the lemma holds for all n. □

By Lemma 6, the higher-order derivatives of log empirical likelihood are rational functions of the sample moments of higher-order derivatives of g. We can anticipate that higher order derivatives of log empirical likelihood can be bounded in a small neighborhood of the true parameter when sample size is large, provided the population moments of higher-order derivatives of function g are finite. This we prove in the following lemma.

Lemma 7. *Under Assumptions 3, 4 and 5, there exist constants δ_2, C_3 and N_4 such that for any $|b(\theta - \tilde{\theta})| \leq \delta_2$ and $n > N_4$, any $j = 1, \ldots, k$,*

$$\left| \frac{d^j \tilde{l}(\theta)}{d\theta^j} \right| \leq C_3. \tag{B.1}$$

Proof. All ω_i in Lemma 6 are equal to 1 when evaluated at $\tilde{\theta}$. Under the assumption of finite moments, by strong law of large numbers, and strong consistency of the M-estimator $\tilde{\theta}$, we have

$$M_j = \frac{1}{n} \sum_{i=1}^{n} \prod_l \frac{d^l g\left(X_i, \tilde{\theta}\right)}{d\theta^l} \rightarrow E\left\{ \prod_l \frac{d^l g\left(X_1, \theta_0\right)}{d\theta^l} \right\} < \infty, \text{ a.s.} .$$

By Lemma 3, the higher order derivatives are continuous functions. Hence for any small number $\varepsilon_2 > 0$, there exists a constant δ_2 such that whenever $|b(\theta - \tilde{\theta})| \leq \delta_2$,

$$\left| \frac{d^j \tilde{l}(\theta)}{d\theta^j} - \frac{d^j \tilde{l}\left(\tilde{\theta}\right)}{d\theta^j} \right| < \varepsilon_2.$$

By Lemma 6, there exists a constant N_4, such that whenever $n > N_4$.

$$\left| \frac{d^j \tilde{l}(\theta)}{d\theta^j} - \frac{P_k\left[E\left\{ \prod_l d^l g\left(X_1, \theta_0\right) / d\theta^l \right\}, \ldots, 0, 1 \right]}{D^{r_k}} \right| < \varepsilon_2.$$

By assumption, all the moments are bounded when $k \leq K + 3$. Then there exist a constant C_3, such that

$$D^{-r_k} P_k\left[E\left\{ \prod_l \frac{d^l g\left(X_1, \theta_0\right)}{d\theta^l} \right\}, \ldots, 0, 1 \right] \leq C_3,$$

which leads to (B.1). $\qquad\square$

Appendix C: Expansion near the M-estimator

Lemma 8. *Under Assumptions 1 and 2, there exists a $\delta_3 > 0$, such that*

$$\sum_{i=1}^{n} \log \hat{w}_i(\theta) - \sum_{i=1}^{n} \log \hat{w}_i\left(\tilde{\theta}\right) \leq -\frac{1}{4} y^2,$$

for any $\theta \in \{\theta : |b(\theta - \tilde{\theta})| < \delta_3\} \cap H$.

Proof. By Taylor expansion,

$$\frac{1}{n} \left\{ \sum_{i=1}^{n} \log \hat{w}_i(\theta) - \sum_{i=1}^{n} \log \hat{w}_i\left(\tilde{\theta}\right) \right\} = \frac{d\tilde{l}\left(\tilde{\theta}\right)}{d\theta}\left(\theta - \tilde{\theta}\right) + \frac{1}{2} \frac{d^2\tilde{l}(\theta^*)}{d\theta^2}\left(\theta - \tilde{\theta}\right)^2,$$

where $|\theta^* - \tilde{\theta}| \le |\theta - \tilde{\theta}|$. By Lemma 4, the first term in above equation is zero. By Lemma 6, we know that $\mathrm{d}^2\tilde{l}(\theta)/\mathrm{d}\theta^2$ is a continuous function in θ. Thus there exists a δ_3, such that for any $|b(\theta^* - \tilde{\theta})| \le |b(\theta - \tilde{\theta})| < \delta_3$,

$$\left| \frac{\mathrm{d}^2\hat{l}(\theta^*)}{\mathrm{d}\theta^2} \left(\theta - \tilde{\theta}\right)^2 + \left| b\left(\theta - \tilde{\theta}\right)\right|^2 \right| < \frac{1}{2} \left| b\left(\theta - \tilde{\theta}\right)\right|^2 .$$

Hence $\mathrm{d}^2\tilde{l}(\theta^*)/\mathrm{d}\theta^2(\theta - \tilde{\theta})^2 < -|b(\theta - \tilde{\theta})|^2/2$, so that,

$$\sum_{i=1}^{n} \log \hat{w}_i\left(\theta\right) - \sum_{i=1}^{n} \log \hat{w}_i\left(\tilde{\theta}\right) < \frac{1}{2} \times \frac{1}{2} \left|\sqrt{n}b\left(\theta - \tilde{\theta}\right)\right|^2 = \frac{1}{4}y^2. \qquad \square$$

The next lemma plays a key role in expanding the posterior, and can be interpreted as an empirical likelihood version of the Edgeworth expansion.

Lemma 9. *Under Assumptions 1, 3, 4 and 5, then there exist δ_4, M_3 and N_5, such that*

$$\left| \int_{-\sqrt{n}\delta_4}^{\sqrt{n}\delta_4} \exp\left\{ -\frac{1}{2}y^2 + \sum_{k=3}^{K+3} a_{kn}\left(\frac{y}{b}\right)^k n^{-(k-2)/2} \right\} - \prod_{i=1}^{n} \frac{\hat{w}_i\left(\theta\right)}{\hat{w}_i\left(\tilde{\theta}\right)} \, \mathrm{d}y \right|$$

$$\le M_3 n^{-(K+2)/2},$$

almost surely.

Proof. Let $\delta_4 \le \min(\delta_2, \delta_3)$ in Lemma 7 and Lemma 8

$$\int_{-\sqrt{n}\delta_4}^{\sqrt{n}\delta_4} \exp\left\{ -\frac{1}{2}y^2 + \sum_{k=3}^{K+3} a_{kn}\left(\frac{y}{b}\right)^k n^{-(k-2)/2} \right\} - \prod_{i=1}^{n} \frac{\hat{w}_i\left(\theta\right)}{\hat{w}_i\left(\tilde{\theta}\right)} \, \mathrm{d}y$$

$$= \int_{-\sqrt{n}\delta_4}^{\sqrt{n}\delta_4} \exp\left\{ \sum_{i=1}^{n} \log \hat{w}_i\left(\theta\right) - \sum_{i=1}^{n} \log \hat{w}_i\left(\tilde{\theta}\right) \right\}$$

$$\left[\exp\left\{ -\frac{1}{2}y^2 + \sum_{k=3}^{K+3} a_{kn}\left(\frac{y}{b}\right)^k n^{-(k-2)/2} - \sum_{i=1}^{n} \log \hat{w}_i\left(\theta\right) \right.\right.$$

$$\left.\left. + \sum_{i=1}^{n} \log \hat{w}_i\left(\tilde{\theta}\right) \right\} - 1 \right] \mathrm{d}y.$$

By Lemma 8, Lemma 7 and Taylor expansion, the above equation is bounded by

$$\int_{-\sqrt{n}\delta_4}^{\sqrt{n}\delta_4} \exp\left(-\frac{y^2}{4} \right) \left| \exp\left\{ -a_{K+4,n}\left(\theta^*\right)\left(\frac{y}{b}\right)^{K+4} n^{-(K+2)/2} \right\} - 1 \right| \mathrm{d}y$$

$$\le \int_{-\sqrt{n}\delta_4}^{\sqrt{n}\delta_4} \exp\left(-\frac{y^2}{4} \right) \left| \exp\left\{ -C_1\left(\frac{y}{b}\right)^{K+4} n^{-(K+2)/2} \right\} - 1 \right| \mathrm{d}y. \qquad \text{(C.1)}$$

where $|\theta^* - \tilde{\theta}| \le |\theta - \tilde{\theta}| < \delta_4$, and C_4 is some constant dependent on δ_4, N_5 and $a_{K+4,n}(\tilde{\theta})$. For sufficiently large n, and sufficiently small δ_4, $a_{K+4,n}(\theta^*)$ is very close to $a_{K+4,n}(\tilde{\theta})$, and by Lemma 7, $a_{K+4,n}(\tilde{\theta})$ is finite. Hence, for very large n, $\exp\{-C_4(y/b)^{K+4}n^{-(k+2)/2}\} - 1$ does not change sign on either $[-\sqrt{n}\delta_4, 0]$ and $[0, \sqrt{n}\delta_4]$. So without loss of generality, we assume $\exp\{-C_4(y/b)^{K+4}n^{-(K+2)/2}\} - 1 \ge 0$ on $[-\sqrt{n}\delta_4, \sqrt{n}\delta_4]$. With $t = \sqrt{n}$, and $t \in \mathbb{R}^+$, the last term in (C.1) can be written as

$$\int_{-\delta_4 t}^{\delta_4 t} \exp\left(-\frac{y^2}{4}\right) \left\{\exp\left(-\frac{C_4}{b^{K+4}}y^{K+4}t^{-K-2}\right) - 1\right\} dy.$$

If we can show that

$$\lim_{t \to +\infty} \frac{\int_{-\delta_4 t}^{\delta_4 t} \exp\left(-y^2/4\right) \left\{\exp\left(-C_4 y^{K+4} t^{-K-2}/b^{K+4}\right) - 1\right\} dy}{t^{-K-2}} = C_5,$$

for some $C_5 < \infty$, the lemma is proved. Take the derivative with respect to t in both the numerator and the denominator. In the denominator, $(t^{-K-2})' = -(K+2)t^{-K-3}$. In the numerator,

$$\frac{d}{dt} \int_{-\delta_4 t}^{\delta_4 t} \exp\left(-\frac{y^2}{4}\right) \left\{\exp\left(-\frac{C_4}{b^{K+4}}y^{K+4}t^{-K-2}\right) - 1\right\} dy$$

$$= \int_{-\delta_4 t}^{\delta_4 t} \exp\left(-\frac{y^2}{4}\right) \left(-\frac{C_4}{b^{K+4}}y^{K+4}\right)(-K-2)t^{-K-3} \exp\left(-\frac{C_4}{b^{K+4}}\frac{y^{K+4}}{t^{K+2}}\right) dy$$

$$+ \exp\left\{-\frac{(\delta_4 t)^2}{4}\right\}\left[\exp\left\{-\frac{C_4}{b^{K+4}}(\delta_4 t)^{K+4}t^{-K-2}\right\} - 1\right]\delta_4 - \exp\left\{-\frac{(-\delta_4 t)^2}{4}\right\}$$

$$\times \left[\exp\left\{-\frac{C_4}{b^{K+4}}(-\delta_4 t)^{K+4}t^{-K-2}\right\} - 1\right](-\delta_4)$$

$$= \frac{(K+2)C_4}{b^{K+4}}t^{-K-3}\int_{-\delta_4 t}^{\delta_4 t} y^{K+4}\exp\left(-\frac{y^2}{4} - \frac{C_4}{b^{K+4}}y^{K+4}t^{-K-2}\right) dy$$

$$+ \delta_4\left[\exp\left\{-\left(\frac{\delta_4^2}{4} - \frac{C_4}{b^{K+4}}\delta_4^{K+4}\right)t^2\right\} - \exp\left(-\frac{\delta_4^2 t^2}{4}\right)\right] \tag{C.2}$$

$$+ \delta_4\left(\exp\left[-\left\{\frac{\delta_4^2}{4} - \frac{C_1}{b^{K+4}}(-\delta_4)^{K+4}\right\}t^2\right] - \exp\left(-\frac{\delta_4^2 t^2}{4}\right)\right).$$

We choose δ_4 sufficiently small such that

$$\delta_4 < \min\left(\sqrt[K+2]{\frac{b^{K+4}}{4|C_4|}}, \delta_2, \delta_3\right). \tag{C.3}$$

Hence,

$$0 < \frac{\delta_4^2}{4} - \frac{|C_4|}{b^{K+4}}\delta_4^{K+4} \le \frac{\delta_4^2}{4} - \frac{C_4}{b^{K+4}}(\pm\delta_4)^{K+4},$$

and

$$\lim_{t \to +\infty} \frac{\delta_4 \left[\exp\left\{ -\left(\delta_4^2/4 - |C_4| \, \delta_2^{K+4}/b^{K+4} \right) t^2 \right\} - \exp\left(-\delta_4^2 t^2/4 \right) \right]}{-(K+2) \, t^{-K-3}} = 0.$$

Hence, the last two terms in (C.2) tend to zero when $t \to +\infty$. Now we consider the ratio

$$\frac{\left\{ (K+2) \, C_4/b^{K+4} \right\} t^{-K-3} \int_{-\delta_4 t}^{\delta_4 t} y^{K+4} \exp\left(-y^2/4 - C_4 y^{K+4} t^{-K-2}/b^{K+4} \right) \, dy}{-(K+2) \, t^{-K-3}}$$

$$= -\frac{C_4}{b^{K+4}} \int_{-\delta_4 t}^{\delta_4 t} y^{K+4} \exp\left(-\frac{y^2}{4} - \frac{C_4}{b^{K+4}} y^{K+4} t^{-K-2} \right) \, dy. \tag{C.4}$$

Since δ_4 satisfies (C.3), (C.4) is bounded by

$$\frac{|C_4|}{b^{K+4}} \int_{-\delta_4 t}^{\delta_4 t} |y|^{K+4} \exp\left\{ -\frac{y^2}{4} - \frac{|C_4|}{b^{K+4}} (\delta_4 t)^{K+2} y^2 t^{-K-2} \right\} \, dy$$

$$= \frac{|C_4|}{b^{K+4}} \int_{-\delta_4 t}^{\delta_4 t} |y|^{K+4} \exp\left\{ -\left(\frac{\delta_4^2}{4} - \frac{|C_4|}{b^{K+4}} \delta_4^{K+4} \right) y^2 \right\} \, dy$$

$$\to \frac{|C_4|}{b^{K+4}} \sqrt{2\pi \left\{ 2\left(\frac{\delta_4^2}{4} - \frac{|C_4|}{b^{K+4}} \delta_4^{K+4} \right) \right\}^{-1}} \left\{ 2\left(\frac{\delta_4^2}{4} - \frac{|C_4|}{b^{K+4}} \delta_4^{K+4} \right) \right\}^{-(K+4)/2}$$

$$\times \frac{2^{(K+4)/2} \Gamma\left\{ (K+4+1)/2 \right\}}{\sqrt{\pi}} < \infty.$$

So by L'Hostiple's rule, we have

$$\lim_{t \to +\infty} \frac{\int_{-\delta_4 t}^{\delta_4 t} \exp\left(-y^2/4 \right) \left\{ \exp\left(-C_4 y^{K+4} t^{-K-2}/b^{K+4} \right) - 1 \right\} \, dy}{t^{-K-2}}$$

$$= \lim_{t \to +\infty} \frac{\left[\int_{-\delta_4 t}^{\delta_4 t} \exp\left(-y^2/4 \right) \left\{ \exp\left(-C_4 y^{K+4} t^{-K-2}/b^{K+4} \right) - 1 \right\} \, dy \right]'}{\left(t^{-K-2} \right)'}$$

$$= C_5 < \infty. \qquad \square$$

Lemma 10. *Under Assumptions 1, 3, 4 and 5, there exists $\delta_4 > 0$, and constants M_4, N_6, such that*

$$\left| \int_{-\sqrt{n}\delta_4}^{\sqrt{n}\delta_4} \left[\exp\left\{ -\frac{1}{2} y^2 + \sum_{k=3}^{K+3} a_{kn} \left(\frac{y}{b} \right)^k n^{-(k-2)/2} \right\} \rho_K(\theta) \right. \right.$$

$$\left. \left. - \prod_{i=1}^{n} \frac{\hat{w}_i(\theta)}{\hat{w}_i(\tilde{\theta})} \rho(\theta) \right] dy \right| \leq M_4 n^{-\frac{1}{2}(K+1)}, \quad \text{a.s.} . \tag{C.5}$$

Proof. Use δ_4 in Lemma 9, and apply Taylor expansion of $\tilde{l}(\theta)$ around $\tilde{\theta}$. Then for any $\tilde{\theta} - \delta_4/b \leq \theta \leq \tilde{\theta} + \delta_4/b$, there exists a θ^* which satisfies $|b(\theta^* - \tilde{\theta})| \leq |b(\theta - \tilde{\theta})|$.

This leads to

$$
\tilde{l}(\theta) = \tilde{l}\left(\tilde{\theta}\right) + \frac{\mathrm{d}\tilde{l}\left(\tilde{\theta}\right)}{\mathrm{d}\theta}\left(\theta - \tilde{\theta}\right) + \frac{1}{2}\frac{\mathrm{d}^2\tilde{l}\left(\tilde{\theta}\right)}{\mathrm{d}\theta^2}\left(\theta - \tilde{\theta}\right)^2 + \sum_{k=3}^{K+3} a_{kn}\left(\tilde{\theta}\right)\left(\theta - \tilde{\theta}\right)^k
$$

$$
+ \frac{1}{(K+4)!}\frac{\mathrm{d}^{K+4}}{\mathrm{d}\theta}\tilde{l}\left(\theta^*\right)\left(\theta - \tilde{\theta}\right)^{K+4}
$$

$$
= \tilde{l}\left(\tilde{\theta}\right) - \frac{1}{2}y^2 n^{-1} + \sum_{k=3}^{K+3} a_{kn}\left(\frac{y}{b}\right)^k n^{-k/2}
$$

$$
+ \frac{1}{(K+4)!}\frac{\mathrm{d}^{K+4}}{\mathrm{d}\theta}\tilde{l}\left(\theta^*\right)\left(\theta - \tilde{\theta}\right)^{K+4}.
$$

Now

$$
\left| \exp\left\{ -\frac{1}{2}y^2 + \sum_{k=3}^{K+3} a_{kn}\left(\frac{y}{b}\right)^k n^{-(k-2)/2} \right\}\rho_K(\theta) - \prod_{i=1}^{n}\frac{\hat{w}_i(\theta)}{\hat{w}_i\left(\tilde{\theta}\right)}\rho(\theta) \right|
$$

$$
\leq \left| \exp\left\{ -\frac{1}{2}y^2 + \sum_{k=3}^{K+3} a_{kn}\left(\frac{y}{b}\right)^k n^{-(k-2)/2} \right\}\rho_K(\theta) - \prod_{i=1}^{n}\frac{\hat{w}_i(\theta)}{\hat{w}_i\left(\tilde{\theta}\right)}\rho_K(\theta) \right|
$$

$$
+ \left| \prod_{i=1}^{n}\frac{\hat{w}_i(\theta)}{\hat{w}_i\left(\tilde{\theta}\right)}\rho_K(\theta) - \prod_{i=1}^{n}\frac{\hat{w}_i(\theta)}{\hat{w}_i\left(\tilde{\theta}\right)}\rho(\theta) \right|
$$

$$
\leq |\rho_K(\theta)|\exp\left\{ n\tilde{l}(\theta) - n\tilde{l}\left(\tilde{\theta}\right) \right\}\left| \exp\left[n\left\{\tilde{l}\left(\tilde{\theta}\right) - \frac{1}{2}y^2 n^{-1} \right. \right. \right.
$$

$$
\left. \left. \left. + \sum_{k=3}^{K+3} a_{kn}\left(\frac{y}{b}\right)^k n^{-k/2} - \tilde{l}(\theta) \right\} \right] - 1 \right|
$$

$$
+ \exp\left\{ n\tilde{l}(\theta) - n\tilde{l}\left(\tilde{\theta}\right) \right\}\left| \rho_K(\theta) - \rho(\theta) \right|.
$$

By Lemma 9, the first term in the right hand side is bounded by

$$
\left\{ \max_{\tilde{\theta}-\delta_4/b \leq \theta \leq \tilde{\theta}+\delta_4/b} \rho_K(\theta) \right\} M_3 n^{-(K+2)/2}.
$$

By Lemma 8, and Taylor expansion of $\rho(\theta)$, the second term is bounded by

$$
\int_{-\sqrt{n}\delta_4}^{\sqrt{n}\delta_4} \exp\left(-\frac{y^2}{4} \right)\frac{n^{-(K+1)/2}}{(K+1)!}\rho^{K+1}(\theta^*) y^{K+1}\, \mathrm{d}y
$$

$$
\leq \frac{n^{-(K+1)/2}}{(K+1)!}\left\{ \max_{\tilde{\theta}-\delta_4/b \leq \theta \leq \tilde{\theta}+\delta_4/b} \rho^{K+1}(\theta^*) \right\}\left\{ \int_{-\sqrt{n}\delta_4}^{\sqrt{n}\delta_4} \exp\left(-\frac{y^2}{4} \right) y^{(K+1)/2}\, \mathrm{d}y \right\}
$$

$$
= \frac{n^{-(K+1)/2}}{(K+1)!}\left\{ \max_{\tilde{\theta}-\delta_4/b \leq \theta \leq \tilde{\theta}+\delta_4/b} \rho^{K+1}(\theta^*) \right\}\int_{0}^{\infty} \exp\left(-\frac{y^2}{4} \right) y^{(K+1)/2}\, \mathrm{d}y
$$

Hence (C.5) holds.   □

Appendix D: Proof of the fundamental theorem for expansion

We first intuitively derive $P_K(\xi, n)$. First, we expand

$$
\exp\left\{\sum_{k=3}^{K+3} a_{kn}\left(\frac{y}{b}\right)^k n^{-(k-2)/2}\right\}
$$

$$
= \sum_{i=0}^{K+1}\frac{1}{i!}\left\{\sum_{k=3}^{K+3} a_{kn}\left(\frac{y}{b}\right)^k n^{-(k-2)/2}\right\}^i
$$

$$
= 1 + \sum_{i=1}^{K+1}\frac{1}{i!}\sum_{\sum_{u=3}^{K+3} m_{u,i}=i}\binom{i}{m_{3,i},\ldots,m_{K+3,i}}\prod_{u=3}^{K+3}(a_{un})^{m_{u,i}}
$$

$$
\times \left(\frac{y}{b}\right)^{\sum_{u=3}^{K+3} m_{u,i}u} n^{-\left\{\sum_{u=3}^{K+3} m_{u,i}(u-2)\right\}/2}.
$$

Multiplying the above expression by ρ_K,

$$
\exp\left\{\sum_{k=3}^{K+3} a_{kn}\left(\frac{y}{b}\right)^k n^{-(k-2)/2}\right\}\rho_K\left(\theta\right)
$$

$$
= \left[1 + \sum_{i=1}^{K+1}\frac{1}{i!}\sum_{\sum_{u=3}^{K+3} m_{u,i}=i}\binom{i}{m_{3,i},\ldots,m_{K+3,i}}\prod_{u=3}^{K+3}(a_{un})^{m_{u,i}}\left(\frac{y}{b}\right)^{\sum_{u=3}^{K+3} m_{u,i}u}\right.
$$

$$
\left.\times n^{-\left\{\sum_{u=3}^{K+3} m_{u,i}(u-2)\right\}/2}\right] \times \left\{\rho\left(\tilde{\theta}\right) + \sum_{j=1}^{K}\frac{1}{j!}\rho^{(j)}\left(\frac{y}{b}\right)^j n^{-j/2}\right\}
$$

$$
= \rho\left(\tilde{\theta}\right) + \sum_{j=1}^{K}\frac{1}{j!}\rho^{(j)}\left(\frac{y}{b}\right)^j n^{-j/2} + \rho\left(\tilde{\theta}\right)\sum_{i=1}^{K+1}\frac{1}{i!}
$$

$$
\times \sum_{\sum_{u=3}^{K+3} m_{u,i}=i}\binom{i}{m_{3,i},\ldots,m_{K+3,i}}\prod_{u=3}^{K+3}(a_{un})^{m_{u,i}}\left(\frac{y}{b}\right)^{\sum_{u=3}^{K+3} m_{u,i}u}
$$

$$
\times n^{-\left\{\sum_{u=3}^{K+3} m_{u,i}(u-2)\right\}/2} + \left[\sum_{i=1}^{K+1}\frac{1}{i!}\sum_{\sum_{u=3}^{K+3} m_{u,i}=i}\binom{i}{m_{3,i},\ldots,m_{K+3,i}}\right.
$$

$$
\left.\times \prod_{u=3}^{K+3}(a_{un})^{m_{u,i}}\left(\frac{y}{b}\right)^{\sum_{u=3}^{K+3} m_{u,i}u} n^{-\left\{\sum_{u=3}^{K+3} m_{u,i}(u-2)\right\}/2}\right]
$$

$$
\times \sum_{j=1}^{K}\frac{1}{j!}\rho^{(j)}\left(\frac{y}{b}\right)^j n^{-j/2}.
$$

For the third term in the right hand side of the above equation, we change the summation index. Let $\sum_{u=3}^{K+3} m_{u,i}(u-2) = h$. For any $\sum_{u=3}^{K+3} m_{u,i} = i$, $i \leq h \leq i(K+1)$, $h/(K+1) \leq i \leq h$. Thus the third term in the summation can be rearranged as

$$
\sum_{h=1}^{(K+1)^2} \left\{ \rho\left(\tilde{\theta}\right) \sum_{i=\lceil h/(K+1)\rceil}^{h} \frac{1}{i!} \sum_{I_{i,h}} \binom{i}{m_{3,i}, \ldots, m_{K+3,i}} \right.
$$
$$
\left. \times \prod_{u=3}^{K+3} (a_{un})^{m_{u,i}} \left(\frac{y}{b}\right)^{\sum_{u=3}^{K+3} m_{u,i} u} \right\} n^{-h/2}.
$$

Similarly for the fourth term, let $\sum_{u=3}^{K+3} m_{u,i}(u-2)+j = h$. Then the summation can be rearranged as

$$
\sum_{h=2}^{(K+1)^2+K} \left\{ \sum_{j=1}^{h-1} \frac{1}{j!} \rho^{(j)} \left(\frac{y}{b}\right)^j \sum_{i=\lceil (h-j)/(K+1)\rceil}^{h-j} \frac{1}{i!} \sum_{I_{i,h-j}} \binom{i}{m_{3,i}, \ldots, m_{K+3,i}} \right.
$$
$$
\left. \times \prod_{u=3}^{K+3} (a_{un})^{m_{u,i}} \left(\frac{y}{b}\right)^{\sum_{u=3}^{K+3} m_{u,i} u} \right\} n^{-h/2}.
$$

We collect same order terms of n, and denote the summation of all the terms with order higher than K to be $R_K(Y)$. Then we get the product as

$$
\rho\left(\tilde{\theta}\right) + \left\{ \rho'\left(\frac{y}{b}\right) + \rho\left(\tilde{\theta}\right) a_{3n} \left(\frac{y}{b}\right)^3 \right\} n^{-1/2}
$$
$$
+ \sum_{h=2}^{K} \left\{ \frac{1}{h!} \rho^{(h)} \left(\frac{y}{b}\right)^h + \sum_{j=0}^{h-1} \frac{1}{j!} \rho^{(j)} \left(\frac{y}{b}\right)^j \right.
$$
$$
\times \sum_{i=\lceil (h-j)/(K+1)\rceil}^{h-j} \frac{1}{i!} \sum_{I_{i,h-j}} \binom{i}{m_{3,i}, \ldots, m_{K+3,i}} \prod_{u=3}^{K+3} (a_{un})^{m_{u,i}} \left(\frac{y}{b}\right)^{\sum_{u=3}^{K+3} m_{u,i} u} \right\}
$$
$$
\times n^{-h/2} + R_K(y).
$$

Integrating any Borel set $(Y_{(1)}, \xi]$, we get the polynomial $P_K(\xi, n)$. Now we prove Theorem 1.

Proof. Let $A_1 = \{|y| \geq \delta_4 \sqrt{n}\} \cap (Y_{(1)}, \xi)$ and $A_2 = \{|y| < \delta_4 \sqrt{n}\} \cap (Y_{(1)}, \xi)$, where δ_4 is in Lemma 10. Then

$$
\left| \int_{Y_{(1)}}^{\xi} \prod_{i=1}^{n} \frac{\hat{w}_i\left(\tilde{\theta} + y/\sqrt{n}b\right)}{\hat{w}_i\left(\tilde{\theta}\right)} \rho\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) \, dy - P_K(\xi, n) \right|
$$
$$
= \left| \int_{Y_{(1)}}^{\xi} \left\{ \prod_{i=1}^{n} \frac{\hat{w}_i\left(\tilde{\theta} + y/\sqrt{n}b\right)}{\hat{w}_i\left(\tilde{\theta}\right)} \rho\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) \right.\right.
$$

$$- \exp\left(-\frac{y^2}{2}\right) \sum_{h=0}^{K} \alpha_h(y, n) \, n^{-h/2}\Bigg\} \mathrm{d}y \Bigg|$$

$$\leq \left| \int_{A_1} \left\{ \prod_{i=1}^{n} \frac{\hat{w}_i\left(\tilde{\theta} + y/\sqrt{n}b\right)}{\hat{w}_i\left(\tilde{\theta}\right)} \rho\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) \right. \right.$$

$$\left. - \exp\left(-\frac{y^2}{2}\right) \sum_{h=0}^{K} \alpha_h(y, n) \, n^{-h/2} \right\} \mathrm{d}y \Bigg|$$

$$+ \left| \int_{A_2} \left\{ \prod_{i=1}^{n} \frac{\hat{w}_i\left(\tilde{\theta} + y/\sqrt{n}b\right)}{\hat{w}_i\left(\tilde{\theta}\right)} \rho\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) \right. \right.$$

$$\left. - \exp\left(-\frac{y^2}{2}\right) \sum_{h=0}^{K} \alpha_h(y, n) \, n^{-h/2} \right\} \mathrm{d}y \Bigg|.$$

For the first term, by Lemma 5, we have

$$\left| \int_{A_1} \prod_{i=1}^{n} \frac{\hat{w}_i\left(\tilde{\theta} + y/\sqrt{n}b\right)}{\hat{w}_i\left(\tilde{\theta}\right)} \rho\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) - \exp\left(-\frac{y^2}{2}\right) \sum_{h=0}^{K} \alpha_h(y, n) \, n^{-h/2} \, \mathrm{d}y \right|$$

$$\leq \int_{A_1} \exp\left\{ n\hat{l}\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) - n\hat{l}\left(\tilde{\theta}\right) \right\} \rho\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) \, \mathrm{d}y$$

$$+ \left| \int_{A_1} \exp\left(-\frac{y^2}{4}\right) \sum_{h=0}^{K} \alpha_h(y, n) \, n^{-h/2} \, \mathrm{d}y \right|$$

$$\leq \exp\left(-n\varepsilon\right) \int_{A_1} \rho\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) \, \mathrm{d}y$$

$$+ \exp\left(-\frac{\delta_4^2 n}{4}\right) \left| \int_{A_1} \exp\left(-\frac{y^2}{4}\right) \sum_{h=0}^{K} \alpha_h(y, n) \, n^{-h/2} \, \mathrm{d}y \right|$$

$$\leq \exp\left(-n\varepsilon\right) \int_{\mathbb{R}} \rho\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) \, \mathrm{d}y + \exp\left(-n\frac{\delta_4^2}{4}\right)$$

$$\times \sum_{h=0}^{K} \left\{ \int_{Y_{(1)}}^{Y_{(n)}} \exp\left(-\frac{y^2}{4}\right) |\alpha_h(y, n)| \, \mathrm{d}y \right\} n^{-h/2}.$$

The above terms are exponentially decreasing with respect to n. Hence there exist N_1, and M_5, such that for any $n \geq N_1$,

$$\left| \int_{A_1} \prod_{i=1}^{n} \frac{\hat{w}_i\left(\tilde{\theta} + y/\sqrt{n}b\right)}{\hat{w}_i\left(\tilde{\theta}\right)} \rho\left(\tilde{\theta} + \frac{y}{\sqrt{n}b}\right) \right.$$

$$\left. - \exp\left(-\frac{y^2}{2}\right) \sum_{h=0}^{K} \alpha_h(y, n) \, n^{-h/2} \, \mathrm{d}y \right| \leq M_5 n^{-(K+1)/2} \qquad .$$

For the second term, by Lemma 10, we have

$$\left| \int_{A_2} \prod_{i=1}^{n} \frac{\hat{w}_i \left(\tilde{\theta} + y/\sqrt{n}b \right)}{\hat{w}_i \left(\tilde{\theta} \right)} \rho \left(\tilde{\theta} + \frac{y}{\sqrt{n}b} \right) - \exp \left(-\frac{y^2}{2} \right) \sum_{h=0}^{K} \alpha_h \left(y, n \right) n^{-h/2} \, dy \right|$$

$$\leq \left| \int_{A_2} \prod_{i=1}^{n} \frac{\hat{w}_i \left(\tilde{\theta} + y/\sqrt{n}b \right)}{\hat{w}_i \left(\tilde{\theta} \right)} \rho \left(\tilde{\theta} + \frac{y}{\sqrt{n}b} \right) \right.$$

$$\left. - \exp \left\{ -\frac{1}{2} y^2 + \sum_{k=3}^{K+3} a_{kn} \left(\tilde{\theta} \right) \left(\frac{y}{b} \right)^k n^{-(K-2)/2} \right\} \rho_K \left(\theta \right) dy \right|$$

$$+ \left| \int_{A_2} \exp \left\{ -\frac{1}{2} y^2 + \sum_{k=3}^{K+3} a_{kn} \left(\tilde{\theta} \right) \left(\frac{y}{b} \right)^k n^{-(K-2)/2} \right\} \rho_K \left(\tilde{\theta} + \frac{y}{\sqrt{n}b} \right) \right.$$

$$\left. - \exp \left(-\frac{y^2}{2} \right) \sum_{h=0}^{K} \alpha_h \left(y, n \right) n^{-h/2} \, dy \right|$$

$$\leq M_4 n^{-(K+1)/2}$$

$$+ \left| \int_{A_2} \exp \left\{ -\frac{1}{2} y^2 + \sum_{k=3}^{K+3} a_{kn} \left(\tilde{\theta} \right) \left(\frac{y}{b} \right)^k n^{-(K-2)/2} \right\} \rho_K \left(\tilde{\theta} + \frac{y}{\sqrt{n}b} \right) \right.$$

$$\left. - \exp \left(-\frac{y^2}{2} \right) \sum_{h=0}^{K} \alpha_h \left(y, n \right) n^{-h/2} \, dy \right|.$$

For the second term in the right hand side, we add and subtract $R_K(y)$ in integrand, and by Taylor expansion,

$$\left| \int_{A_2} \exp \left\{ -\frac{1}{2} y^2 + \sum_{k=3}^{K+3} a_{kn} \left(\tilde{\theta} \right) \left(\frac{y}{b} \right)^k n^{-(K-2)/2} \right\} \rho_K \left(\tilde{\theta} + \frac{y}{\sqrt{n}b} \right) \right.$$

$$\left. - \exp \left(-\frac{y^2}{2} \right) \sum_{h=0}^{K} \alpha_h \left(y, n \right) n^{-h/2} \, dy \right|$$

$$\leq \left| \int_{A_2} \exp \left(-\frac{y^2}{2} \right) \rho_K \left(\theta \right) \left[\exp \left\{ \sum_{k=3}^{K+3} a_{kn} \left(\tilde{\theta} \right) \left(\frac{y}{b} \right)^k n^{-(K-2)/2} \right\} \right. \right.$$

$$\left. \left. - \sum_{i=0}^{K+1} \frac{1}{i!} \left\{ \sum_{k=3}^{K+3} a_{kn} \left(\tilde{\theta} \right) \left(\frac{y}{b} \right)^k n^{-(K-2)/2} \right\}^i \right] dy \right|$$

$$+ \left| \int_{A_2} \exp \left(-\frac{y^2}{2} \right) R_K \left(y \right) \, dn^{-1/2} y \right|$$

$$= \left| \int_{A_2} \exp \left(-\frac{y^2}{2} \right) \rho_K \left(\theta \right) \frac{1}{(K+2)!} \exp \left(L \right) \right.$$

$$\left\{ \sum_{k=3}^{K+3} a_{kn}\left(\tilde{\theta}\right)\left(\frac{y}{b}\right)^k n^{-(K-2)/2}\right\}^{K+2} dy \right| + \left| \int_{A_2} \exp\left(-\frac{y^2}{2}\right) R_K\left(y\right) dy \right|,$$

where $|L| \le |\sum_{k=3}^{K+3} a_{kn}(\tilde{\theta})(y/b)^k n^{-(K-2)/2}|$. We know that $R_K(y)$ is a polynomial with order $n^{-(K+1)/2}$. So there exists an M_6, such that

$$\left| \int_{A_2} \exp\left(-\frac{y^2}{2}\right) R_K\left(y\right) dy \right| \le M_6 n^{-(K+1)/2}.$$

For the first term,

$$
\left| \int_{A_2} \exp\left(-\frac{y^2}{2}\right) \rho_K\left(\theta\right) \frac{1}{(K+2)!} \exp\left(L\right) \right.
$$

$$
\left. \times \left\{ \sum_{k=3}^{K+3} a_{kn}\left(\tilde{\theta}\right)\left(\frac{y}{b}\right)^k n^{-(K-2)/2}\right\}^{K+2} dn^{-1/2}Y \right| \tag{D.1}
$$

$$
\le \frac{1}{(K+2)!} \left| \int_{A_2} \exp\left(-\frac{y^2}{2}\right) \rho_K\left(\theta\right) \exp\left\{ \left| \sum_{k=3}^{K+3} a_{kn}\left(\tilde{\theta}\right)\left(\frac{y}{b}\right)^k n^{-(K-2)/2}\right| \right\} \right.
$$

$$
\left. \times \left\{ \sum_{k=3}^{K+3} a_{kn}\left(\tilde{\theta}\right)\left(\frac{y}{b}\right)^k n^{-(K-2)/2}\right\}^{K+2} dy \right|
$$

$$
= \frac{1}{(K+2)!} \left| \int_{A_2} \rho_K\left(\theta\right) \exp\left\{ -\frac{y^2}{2} + \left| \sum_{k=3}^{K+3} a_{kn}\left(\tilde{\theta}\right)\left(\frac{y}{b}\right)^k n^{-(K-2)/2}\right| \right\} \right.
$$

$$
\left. \times \left\{ \sum_{k=3}^{K+3} a_{kn}\left(\tilde{\theta}\right)\left(\frac{y}{b}\right)^k n^{-(K-2)/2}\right\}^{K+2} dy \right|.
$$

We need δ_4 sufficiently small, so that there exist C_6 and C_7, such that

$$
\frac{y^2}{2} - \left| \sum_{k=3}^{K+3} a_{kn}\left(\tilde{\theta}\right)\left(\frac{y}{b}\right)^k n^{-(K-2)/2}\right| \ge C_6 y^2,
$$

$$
\sum_{k=3}^{K+3} a_{kn}\left(\tilde{\theta}\right)\left(\frac{y}{b}\right)^k n^{-(K-2)/2} \le C_7 y^3 n^{-1/2}.
$$

Hence, (D.1) is bounded by

$$
\frac{n^{K+2}}{(K+2)!} \left| \int_{A_n} \exp\left(-C_6 y^2\right)\left(C_7 y^3 n^{-1/2}\right)^{K+2} dy \right|
$$

$$
\le \frac{C_7^{K+2}}{(K+2)!} \left| \int_{Y_{(1)}}^{Y_{(n)}} \exp\left(-C_6 y^2\right) y^{3(K+2)} n^{-(K+2)/2} dy \right|
$$

$$\leq \quad \frac{C_7^{K+2}}{(K+2)!} \left| \int_{Y_{(1)}}^{Y_{(n)}} \exp\left(-C_6 y^2\right) y^{3(K+2)} \, \mathrm{d}y \right| n^{-(K+2)/2}.$$

Adding all the parts, we get the inequality in Theorem 1 .

□

Acknowledgement

References

[1] Baggerly, K. A., 1998. Empirical likelihood as a goodness-of-fit measure. Biometrika 85 (3), 535–547. MR1665869

[2] Clarke, B., Yuan, A., 2010. Reference priors for empirical likelihoods. in Frontiers of Statistical Decision Making and Bayesian Analysis, 56–68.

[3] Cornuet, J.-M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J.-M., Balding, D. J., Guillemaud, T., Estoup, A., 2008. Inferring population history with diy abc: A user-friendly approach to approximate bayesian computation. Bioinformatics 24 (23), 2713–2719.

[4] Fang, K.-T., Mukerjee, R., 2005. Expected lengths of confidence intervals based on empirical discrepancy statistics. Biometrika 92 (2), 499–503. MR2201375

[5] Fang, K.-T., Mukerjee, R., 2006. Empirical-type likelihoods allowing posterior credible sets with frequentist validity: Higher-order asymptotics. Biometrika 93 (3), 723–733. MR2261453

[6] Ghosal, S., 2000. Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. Journal of Multivariate Analysis 74 (1), 49–68. MR1790613

[7] Ghosal, S., et al., 1999. Asymptotic normality of posterior distributions in high-dimensional linear models. Bernoulli 5 (2), 315–331. MR1681701

[8] Ghosh, J. K., Sinha, B., Joshi, S., 1982. Expansions for posterior probability and integrated bayes risk. Statistical Decision Theory and Related Topics III 1, 403–456. MR0705299

[9] Ghosh, M., Liu, R., 2011. Moment matching priors. Sankhya A 73 (2), 185–201. MR2887093

[10] Godambe, V. P., 1960. An optimum property of regular maximum likelihood estimation. The Annals of Mathematical Statistics, 1208–1211. MR0123385

[11] Hartley, H. O., Rao, J., 1968. A new estimation theory for sample surveys. Biometrika 55 (3), 547–557.

[12] Johnson, R. A., 1970. Asymptotic expansions associated with posterior distributions. The Annals of Mathematical Statistics 41 (3), 851–864. MR0263198

[13] Lazar, N. A., 2003. Bayesian empirical likelihood. Biometrika 90 (2), 319–326. MR1986649

[14] Owen, A. B., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75 (2), 237–249. MR0946049

[15] Owen, A. B., 2010. Empirical likelihood. CRC press.

[16] Rubin, D. B., et al., 1981. The bayesian bootstrap. The Annals of Statistics 9 (1), 130–134. MR0600538

[17] Schennach, S. M., 2005. Bayesian exponentially tilted empirical likelihood. Biometrika 92 (1), 31–46. MR2158608

[18] Schennach, S. M., et al., 2007. Point estimation with exponentially tilted empirical likelihood. The Annals of Statistics 35 (2), 634–672. MR2336862

[19] Thomas, D. R., Grunkemeier, G. L., 1975. Confidence interval estimation of survival probabilities for censored data. Journal of the American Statistical Association 70 (352), 865–871. MR0405766

[20] Vexler, A., Kim, Y. M., Yu, J., Lazar, N. A., Hutson, A. D., 2014. Computing critical values of exact tests by incorporating monte carlo simulations combined with statistical tables. Scandinavian Journal of Statistics 41 (4), 1013–1030. MR3277035

[21] Vexler, A., Tao, G., Hutson, A., 2014. Posterior expectation based on empirical likelihoods. Biometrika 101 (3), 711–718. MR3254911