Higher Order Properties of Bayesian Empirical Likelihood

Xiaolong Zhong and Malay Ghosh
University of Florida

June 22, 2016

Outline

- Introduction
- The Basic Settings
- A Few Lemmas
- Bayesian Asymptotics
- Asymptotic Expansion of Posteriors Based on Empirical Likelihood
- Matching Priors
- Simulation
- Summary and Discussion

Introduction

- Empirical likelihood has been studied quite extensively in the frequentist literature, but the corresponding Bayesian literature is somewhat sparse.
- Bayesian methods hold promise, however, either with subjective or with non-subjective priors.
- In addition, Bayesian methods very often overcome the curse of dimensionality by providing suitable dimension reduction.
- Our goal: 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.
- Bernstein–von Mises theorem comes as a special case.
- Our approach also aids in finding non-subjective priors based on empirical likelihood and its variations as mentioned above.

- Some earlier work: Lazar (2003): Bayesian empirical likelihood; Scennach (2005, 2007): exponentially tilted empirical likelihood.
- Baggerly (1998) 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
- Baggerly as well as Owen (2010) 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.

The Basic Settings

- $X_1, \ldots, X_n$: iid rv satisfying $E[g(X_1, \theta)] = 0$, where $\theta \in \mathcal{R}$.
- The EL is a nonparametric likelihood of the form $\prod_{i=1}^{n} w_i(\theta)$, where w_i is the probability mass assigned to $X_i : i = 1, \ldots, n$.
- The constraints are (i) $w_i > 0$ for all $i = 1, \ldots, n$, (ii) $\sum_{i=1}^{n} w_i = 1$ and (iii) $\sum_{i=1}^{n} w_i g(X_i, \theta) = 0$.
- A simple example: $g(X_i, \theta) = X_i - \theta$, $i = 1, \ldots, n$.
- Domain of θ is $H_n = ([\bigcap_{i=1}^{n}\{g(X_i, \theta) \geq 0\}] \bigcup [\bigcap_{i=1}^{n}\{g(X_i, \theta) \leq 0\}])^c$.
- 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 above constraints.
- The solution : $\hat{w}_i^{\text{EL}} = [n(1 + \nu g(X_i, \theta)]^{-1}$;
- ν (the lagrangian multiplier) satisfies $\sum_{i=1}^{n} g(X_i, \theta)[1 + \nu g(X_i, \theta)]^{-1} = 0$.

- Exponentially titled empirical likelihood : maximize instead $-\sum_{i=1}^{n} w_i \log w_i$ with the same constraints.
- The resulting solution is

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

- ν, the Lagrange multiplier, satisfies

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

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

- The general Cressie–Read divergence criterion given by

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

- We focus on the cases $\lambda \geq 0$ and $\lambda \leq -1$, because in these cases $\mathrm{CR}(\lambda)$ is a convex function of the w_i.
- The limiting cases $\lambda \to 0$ and $\lambda \to -1$ lead to the usual empirical likelihood and the exponentially tilted empirical likelihood.
- $\hat{w}_i^{\mathrm{CR}}(\theta) = \frac{1}{n}\{\mu + \nu g(X_i, \theta)\}^{-1/(\lambda+1)}$.
- $\sum_{i=1}^{n}\{\mu + \nu g(X_i, \theta)\}^{-1/(\lambda+1)} = n$;
- $\sum_{i=1}^{n}\{\mu + \nu g(X_i, \theta)\}^{-1/(\lambda+1)} X_i = 0$.
- The posterior under the prior $\rho(\theta)$ is

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

A Few Lemmas

- Lemma 1. Assume $g(\cdot, \cdot)$ is a continuous function. Then the natural domain H_n is a compact set and is nondecreasing with respect to the sample size n.
- Assunption 1. For any θ in natural domain H_n, and $n \geq 3$, $P\{g(X_i, \theta) = 0\} = 0$, $i = 1, \ldots, n$.
- Assumption 2. $g(x, \theta)$ is a continuous function of θ with a continuous first derivative in θ.
- Lemma 2. Under Assumptions 1 and 2, the ν functions in both empirical likelihood and exponentially tilted empirical likelihood and both μ and ν functions in the Cressie-Read empirical likelihood are smooth functions of θ.

- Assumption 3. $g(x, \theta)$ admits $K+4$ th order derivative in θ.
- Under Assumptions 1 and 3, derivatives of order up to K are smooth functions of θ.
- Define $\tilde{\theta}$ as the solution of $\sum_{i=1}^{n} g(X_i, \theta) = 0$.
- Define $\tilde{l}(\theta) = n^{-1} \sum_{i=1}^{n} \log \hat{w}_i(\theta)$, where $\hat{w}_i(\theta)$ is either $\hat{w}_i^{\text{EL}}$ or $\hat{w}_i^{\text{ET}}$ or $\hat{w}_i^{\text{CR}}$.
- Lemma 4. $d^2 \tilde{l}(\theta)/d\theta^2$ when evaluated at $\theta = \tilde{\theta}$ is less than zero.

Bayesian Asymptotics

- An asymptotic expansion of the posterior density (Johnson, 1970).
- Let $X_1, \ldots, X_n|\theta$ iid with common pdf $f(X|\theta)$, and let $\hat{\theta}_n$ denote the MLE of θ.
- $L_n(\theta) = \prod_1^n f(X_i|\theta)$ and $\ell_n(\theta) = \log L_n(\theta)$.
- $a_i = n^{-1}[d^i \ell_n(\theta)/d\theta^i]_{\theta=\hat{\theta}_n}$, $i = 1, 2, \ldots$.
- $\hat{I}_n == -a_2$, the observed per unit Fisher information number.
- Twice differentiable prior π.
- Let $T_n = \sqrt{n}(\theta - \hat{\theta}_n)\hat{I}_n^{1/2}$, and let $\pi_n^*(t)$ be the posterior pdf of T_n.

- Theorem 1. $\pi_n^*(t) = \phi(t)[1 + n^{-1/2}\gamma_1(t; X_1, \ldots, X_n) + n^{-1}\gamma_2(t; X_1, \ldots, X_n)] + O_p(n^{-3/2})$,
  where $\phi(t)$ is the standard normal pdf.
- $\gamma_1(t; X_1, \ldots, X_n) = \frac{a_3 t^3}{6\hat{I}_n^{3/2}} + \frac{t}{\hat{I}_n^{1/2}} \frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)}$.

$$\begin{aligned}
\gamma_2(t; X_1, \ldots, X_n) &= \frac{a_4 t^4}{24\hat{I}_n^2} + \frac{a_3^2 t^6}{72\hat{I}_n^3} \\
&+ \frac{t^2}{2\hat{I}_n} \frac{\pi''(\hat{\theta}_n)}{\pi(\hat{\theta}_n)} + \frac{a_3 t^4}{6\hat{I}_n^2} \frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)} \\
&- \frac{a_4}{8\hat{I}_n^2} - \frac{15a_3^2}{72\hat{I}_n^3} - \frac{1}{2\hat{I}_n} \frac{\pi''(\hat{\theta}_n)}{\pi(\hat{\theta}_n)} \\
&- \frac{a_3}{2\hat{I}_n^2} \frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)}.
\end{aligned}$$

- Special Case: Bernstein von-Mises Theorem
  $\pi_n^*(t) = \phi(t)[1 + O_p(n^{-1/2})]$.

- Outline of Proof: Write $\pi(\theta|X_1,\ldots,X_n) = \frac{\exp[\ell_n(\theta)]\pi(\theta)}{\int \exp[\ell_n(\theta)]\pi(\theta)d\theta} = \frac{\exp[\ell_n(\theta)-\ell_n(\hat{\theta}_n)]\pi(\theta)}{\int \exp[\ell_n(\theta)-\ell_n(\hat{\theta}_n)]\pi(\theta)d\theta}$.
- Then $T_n = \sqrt{n}(\theta - \hat{\theta}_n)\hat{I}_n^{1/2}$ has pdf $\pi_n^*(t) = C_n^{-1}\exp[\ell_n\{\hat{\theta}_n + t(n\hat{I}_n)^{-1/2}\} - \ell_n(\hat{\theta}_n)]\pi\{\hat{\theta}_n + t(n\hat{I}_n)^{-1/2}\}$, where C_n^{-1} is the normalizing constant.
- Use the fourth order Taylor expansion of $\ell_n\{\hat{\theta}_n + t(n\hat{I}_n)^{-1/2}\}$ around $\hat{\theta}_n$ and note $\ell_n'(\hat{\theta}_n) = 0$.
- A similar second order Taylor expansion of $\pi\{\hat{\theta}_n + t(n\hat{I}_n)^{-1/2}\}$ around $\hat{\theta}_n$.

Asymptotic Expansion of Posteriors Based on Empirical Likelihood

- Notations: Prior: $\rho(\theta)$. $\tilde{\theta}$ is the solution of $\sum_{i=1}^{n} g(X_i, \theta) = 0$.
- $\rho_1(\theta) = \rho(\tilde{\theta}) + (\theta - \tilde{\theta})\rho'(\tilde{\theta})$.
- $a_{3n}(\theta) = \frac{1}{6}\sum_{i=1}^{n} \frac{d^3\tilde{l}(\theta)}{d\theta^3}$.
- $b(\theta) =$
$\left[\left\{n^{-1}\sum_{i=1}^{n} \partial g(X_i, \theta)/\partial\theta\right\}^2 / \left\{n^{-1}\sum_{i=1}^{n} g^2(X_i, \theta)\right\}\right]^{1/2}$.
- $y = n^{1/2} b(\theta - \tilde{\theta})$.
- $\alpha_0(y, n) = \rho(\tilde{\theta})$.
- $\alpha_1(y, n) = \rho'(\tilde{\theta})y/b + \rho(\tilde{\theta})a_{3n}(\tilde{\theta})(y/b)^3$.

- Let $H_n = [h_1, h_2]$ be the support of θ.
- Define $Y_{(1)} = n^{1/2} b(h_1 - \tilde{\theta})$ and $Y_{(n)} = n^{1/2} b(h_2 - \tilde{\theta})$ as the normalized lower and upper bounds of the support of the distribution.
- $\alpha_k(y, n)$: some long expression which is a kth degree polynomial in y.
- For any $\xi \in (Y_{(1)}, Y_{(n)})$, and $H_n = [h_1, h_2]$, let $P_K(\xi, n) = \sum_{k=0}^{K} \{\int_{Y_{(1)}}^{\xi} \alpha_k(y, n) \exp(-\frac{y^2}{2}) dy\} n^{-k/2}$.
- Assumption 4. For any $l_i \subset \{2, \ldots, 4\}$, $E\{\prod_{i=1}^{j} \frac{d^{l_i} g(X_1, \theta)}{d\theta^{l_i}}\} < \infty$.
- Assumption 5. The M-estimator $\tilde{\theta}$ is a consistent estimator of θ.

- Theorem 1 (Fundamental Expansion Theorem).
  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 second order continuous derivative. Under Assumptions 1-5, there exist constants $N_1 > 0$ and $M_1 > 0$, such that

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

  for any $n > N_1$ and $\xi \in (Y_{(1)}, Y_{(n)})$.
- This theorem can not only be used to prove asymptotic expansion of the posterior cumulative distribution function , but it can also be used to find the asymptotic expansions of the posterior mean, quantiles and many other quantities of interest.

- Next expansion of posterior cdf.
- Posterior cdf:

$$\Pi(\theta \leq \tilde{\theta} + \xi/(nb)^{1/2} | X_1, \ldots, X_n) = \frac{\int_{h_1}^{\tilde{\theta}+\xi/n^{1/2}b} \prod_{i=1}^{n} \tilde{w}_i(\theta)\rho(\theta)d\theta}{\int_{h_1}^{h_2} \prod_{i=1}^{n} \tilde{w}_i(\theta)\rho(\theta)d\theta}.$$

- Notations: $R_n = \left(Y_{(1)}, Y_{(n)}\right)$; $\Phi\left(\xi | R_n\right) = \frac{\int_{Y_{(1)}}^{\xi} \phi(y)dy}{\int_{Y_{(1)}}^{Y_{(n)}} \phi(y)dy}$.
- Define polynomial $\gamma_i\left(\xi, n\right),\ i = 1, \ldots, n$ recursively as
$\int_{Y_{(1)}}^{\xi} \alpha_k\left(y, n\right) \exp\left(-\frac{y^2}{2}\right) dy =$
$\sum_{j=0}^{k} \left\{ \int_{Y_{(1)}}^{Y_{(n)}} \alpha_j\left(y, n\right) \exp\left(-\frac{y^2}{2}\right) dy \right\} \gamma_{k-j}\left(\xi, n\right)$.
- $\gamma_0(\xi, n) = \frac{\Phi(\xi) - \Phi(Y_{(1)})}{\Phi(Y_{(n)}) - \Phi(Y_{(1)})}$.

- Theorem 2. Under the same assumptions as in Theorem 1, there exist constants N_2 and M_2 such that

$$|\Pi\left(\theta \le \tilde{\theta} + \frac{\xi}{\sqrt{n}b}|X_1,\ldots,X_n\right) - \sum_{i=0}^{K} \gamma_i(\xi,n)\, n^{-i/2}| \le M_2 n^{-(K+1)/2}.$$

- (Bernstein-von Mises Theorem) Under the assumptions of Theorem 1 with $K = 2$, the standardizd posterior distribution function converges to the $N(0,1)$ distribution function almost surely, that is $n^{1/2}b(\theta - \tilde{\theta})|X_1,\ldots,X_n \to \mathrm{N}(0,1)$ a.s.

Matching Priors

- Let $X_1, \ldots, X_n|\theta$ be iid with common pdf $f(X|\theta)$. For $0 < \alpha < 1$, let $\theta^{\pi}_{1-\alpha}(X_1, \ldots, X_n) \equiv \theta^{\pi}_{1-\alpha}$ denote the $(1-\alpha)$th asymptotic posterior quantile of θ based on the prior π, i.e. $P^{\pi}[\theta \leq \theta^{\pi}_{1-\alpha}|X_1, \ldots, X_n] = 1 - \alpha + O_p(n^{-p})$, for some $p > 0$.
- If $P[\theta \leq \theta^{\pi}_{1-\alpha}|\theta] = 1 - \alpha + O_p(n^{-p})$, then some order of probability matching is achieved.
- If $p = 1$, then we call π a first order probability matching prior. If $p = 3/2$, then we call π a second order probability matching prior.

- An intuitive argument why Jeffrey's prior is a first order probability matching prior in the absence of nuisance parameters.
- If $X_1, \ldots, X_n|\theta$ iid $N(\theta, 1)$ and $\pi(\theta) = 1$, $-\infty < \theta < \infty$, then the posterior $\pi(\theta|X_1, \ldots, X_n)$ is $N(\bar{X}_n, n^{-1})$.
- Writing $z_{1-\alpha}$ as the $100(1-\alpha)\%$ quantile of the $N(0,1)$ distribution, one gets
$P[\sqrt{n}(\theta - \bar{X}_n) \leq z_{1-\alpha}|X_1 \ldots, X_n) = 1 - \alpha = P[\sqrt{n}(\bar{X}_n - \theta) \geq -z_{1-\alpha}|\theta]$,
so that the one sided credible interval $\bar{X}_n + z_{1-\alpha}/\sqrt{n}$ has exact frequentist coverage probability $1 - \alpha$.

- The above exact matching does not always hold. However, if $X_1, \ldots, X_n|\theta$ are iid, then $\hat{\theta}_n|\theta$ is asymptotically $N(\theta, (nI(\theta))^{-1})$.
- Then by the delta method $g(\hat{\theta}_n)|\theta \sim N[g(\theta), (g'(\theta))^2(nI(\theta))^{-1}]$.
- So if $g'(\theta) = I^{1/2}(\theta)$ so that $g(\theta) = \int^{\theta} I^{1/2}(t)dt$, one gets $\sqrt{n}[g(\hat{\theta}_n) - g(\theta)]|\theta$ is asymptotically $N(0,1)$.
- Hence, with the uniform prior $\pi(\phi) = 1$ for $\phi = g(\theta)$, coverage matching is asymptotically achieved for ϕ.
- This leads to the prior $\pi(\theta) = \frac{d\phi}{d\theta} = g'(\theta) = I^{1/2}(\theta)$.

- Moment Matching Priors : Let $X_1, \ldots, X_n$ be iid with common pdf $f(X|\theta)$, and $T_n = \sqrt{n}(\theta - \hat{\theta}_n)$.
- The posterior $\pi_n^*(t)$ of T_n is $\pi_n^*(t) = \phi(t)[1 + \frac{1}{\sqrt{n}}(\frac{a_3 t^3}{6\hat{I}_n^{3/2}} + \frac{t}{\hat{I}_n^{1/2}}\frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)}) + O_p(n^{-1})]$.
- $E(\theta|X_1, \ldots, X_n) = \hat{\theta}_n + \frac{1}{n}(\frac{a_3}{2\hat{I}_n^2} + \frac{1}{\hat{I}_n}\frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)}) + O_p(n^{-2})$.
- $n[E(\theta|X_1, \ldots, X_n) - \hat{\theta}_n] \xrightarrow{P} \frac{g_3(\theta)}{2I^2(\theta)} + \frac{1}{I(\theta)}\frac{\pi'(\theta)}{\pi(\theta)}$.
- $g_3(\theta) = E_\theta[\frac{d^3 \log f(X|\theta)}{d\theta^3}]$.
- Moment matching prior: $\pi(\theta) = \exp[-\frac{1}{2}\int^\theta \frac{g_3(v)}{I(v)} dv]$.

- Asymptotic expansion of the posterior mean:
$E\left\{\sqrt{n}b\left(\theta-\tilde{\theta}\right)|X\right\}=$
$\left\{\frac{\rho'(\tilde{\theta})}{\rho(\tilde{\theta})b}\frac{\int_{Y_{(1)}}^{Y_{(n)}} y^2\phi(y)dy}{\int_{Y_{(1)}}^{Y_{(n)}} \phi(y)dy}+\frac{a_{3n}}{b^3}\frac{\int_{Y_{(1)}}^{Y_{(n)}} y^4\phi(y)dy}{\int_{Y_{(1)}}^{Y_{(n)}} \phi(y)dy}\right\}n^{-1}+O_p\left(n^{-\frac{3}{2}}\right).$
- Note $Y_{(n)}\to+\infty$ and $Y_{(1)}\to-\infty$ a.s. as $n\to\infty$.
- $\lim_{n\to\infty}\frac{\int_{Y_{(1)}}^{Y_{(n)}} y^2\phi(y)dy}{\int_{Y_{(1)}}^{Y_{(n)}} \phi(y)dy}=\int_{\mathbb{R}} y^2\phi(y)\,dy=1.$
- $\lim_{n\to\infty}\frac{\int_{Y_{(1)}}^{Y_{(n)}} y^4\phi(y)dy}{\int_{Y_{(1)}}^{Y_{(n)}} \phi(y)dy}=\int_{\mathbb{R}} y^4\phi(y)\,dy=3.$
- The moment matching prior $\rho(\theta)$ is a solution of
$$\frac{\rho'(\theta)}{\rho(\theta)}=-\lim_{n\to\infty}\frac{3a_{3n}}{b^2}$$
- Special Case: $g(x,\theta)=x-\theta$. Then
$\rho(\theta)=\exp\left(\int_{-\infty}^{\theta}\frac{E\{(X_1-s)^3\}}{[E\{(X_1-s)^2\}]^4}ds\right).$

- Estimating the quantile with the aid of posterior mean.
- Begin with the smooth estimating equation $K(x) = \int_{-\infty}^{x} k(u)du$: a kernel smoothing function and $K_h(\cdot) = hK(\cdot/h)$.
- The smoothed constraint of the empirical likelihoods to estimate the α quantile θ_i is $\sum_{i=1}^{n} w_i \left[K_h\left(\theta - X_i\right) - (1-\alpha)\right] = 0$.
- $g\left(\theta, X_i\right) = K_h\left(\theta - X_i\right) - (1-\alpha)$.
- The asymptotic variance is $b^{-2} = \left[\frac{\sum_{i=1}^{n} k_h(\tilde{\theta}-X_i)}{\sum_{i=1}^{n}[K_h\left(\tilde{\theta}-X_i\right)-(1-\alpha)]^2}\right]^{-1}$.
- Prior for θ: $\mathrm{N}(\mu_0, \sigma_0^2)$
- $\frac{\rho'(\theta)}{\rho(\theta)}\Big|_{\theta=\tilde{\theta}} = -\frac{\tilde{\theta}-\mu_0}{\sigma_0^2}$.

Simulations

- Let $g(X_i, \theta) = X_i - \theta,\ i = 1, \ldots, n$ and $K = 3$.
- We compare the first order approximation with normal approximation and second order approximation.
- For all three versions of empirical likelihood, $\tilde{l}^{(3)}\left(\overline{X}\right) = \frac{2n^2 \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^3}{\left\{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right\}^3}$.
- So $\tilde{l}^{(3)}\left(\bar{X}\right)$ 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 $\Phi\left(\xi | R_n\right)$.
- Another complex expression for the second order approximation polynomial.
- 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 $\lambda = 2$.

- The red line: normal approximation of the posterior cdf.
- The blue line: first order approximation of the posterior pdf.
- The green line: the posterior based on the empirical likelihood.
- The purple line: The posterior based on the exponentially tilted empirical likelihood.
- The black line: the Cressie–Read divergence empirical likelihood.
- Even when $n = 10$, the three types of empirical likelihoods are quite close to each other.
- The first order approximation is closer than the normal approximation,
- When the sample size increases to 80, all the lines almost coincide with each other.

- Next 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 use one dimensional constraint $g(X, \theta) = X - \theta$ to estimate the mean.
- The priors are Cauchy distributions with different locations μ_0 and different scales σ_0.
- The accuracy is defined as $E_X \left(\max_y |P(\theta \leq y|X) - \tilde{P}(\theta \leq y|X)| \right)$.
- This measures the performance of our approximations $\tilde{P}(\theta \leq y|X)$ with respect to true Bayesian posteriors.
- The results are summarized in the following table.

- The two columns under Empirical Likelihood means the true posteriors are based on the empirical likelihood.
- 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 documents the simulated accuracies when we use first order approximations to estimate the true posteriors based on the empirical likelihood.
- Other columns need to be interpreted similarly.

Comparison with Parametric Bayesian Model

- The asymptotic variance of the Bayesian empirical likelihood is the inverse of observed Godambe information number.
- In parametric Bayesian model, the, asymptotic variance is the inverse of the observed 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.
- But the EL is more robust against a misspecified model.
- We illustrate this with an example.

- $X_i \overset{iid}{\sim}$ exponential with rate parameter λ.
- Prior for λ: Gamma$(2, 1)$.
- The posterior is then Gamma$(n + 2, n\bar{X} + 1)$.
- The Bayes risk under the true model is $\frac{1+2n}{n(n+1)}$.
- Misspecified model: the data are drawn from χ^2_ν.
- Prior for ν: IG$(2, 1)$.
- Posterior for ν: Inverse Gaussian with mean $\{2/n\log 2\}^{1/2}$ and scale 2.
- The next table shows Bayes risks under the misspecified model for the parametric Bayes and the three versions of EL.

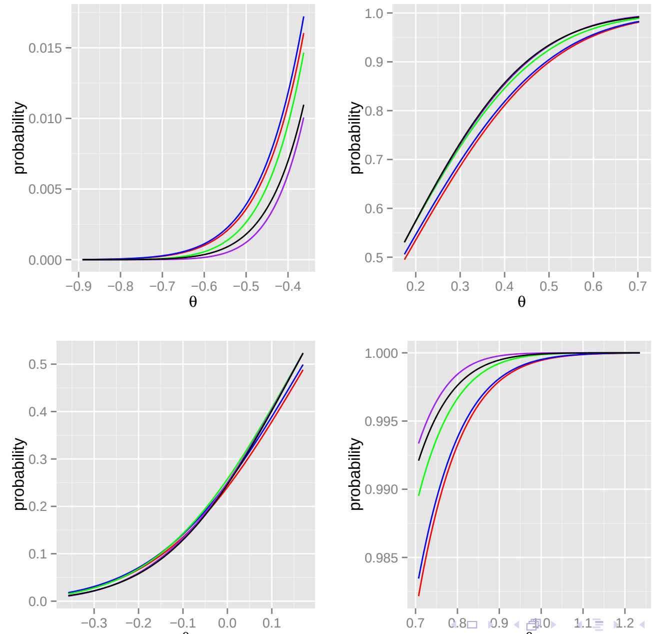
Table 2. Comparison of Bayes Risks

	EL	ETEL	CREL	True	Misspecified
n=10	0.160	0.156	0.159	0.059	0.750
n=20	0.091	0.088	0.090	0.016	0.786
n=50	0.037	0.036	0.036	0.003	0.829
n=100	0.019	0.019	0.019	0.001	0.857

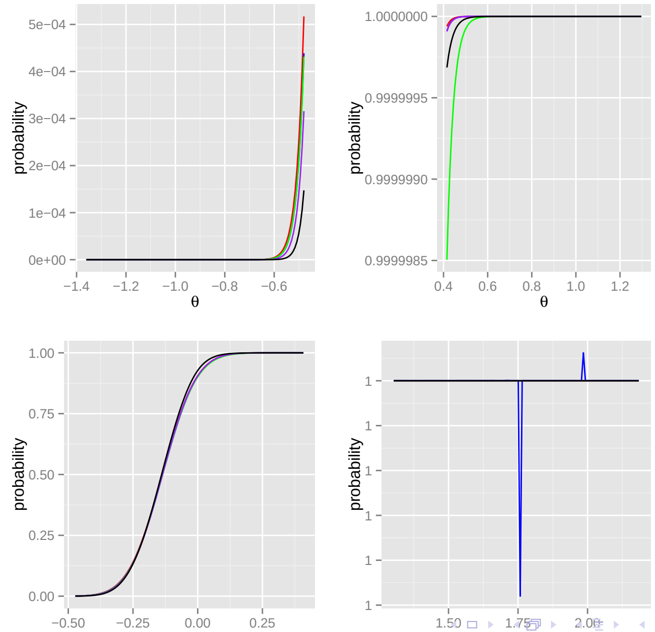
Summary and Future Work

- he 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.
- Current work is extension to the multivariate case as well as expansions subject to multiple constraints.
- A potential topic of research is asymptotic expansion of posteriors under regression constraints.

Posterior Cumulative Distribution Functions When Sample Size is 10



Posterior Cumulative Distribution Functions When Sample Size is 80



Accuracy of Approximation under Different Priors and Different Sample Sizes

n	μ_0	σ_0	EL		ETEL		CR	
			Normal	1st Order	Normal	1st Order	Normal	1st Order
10	1	0.3	0.134	0.070	0.115	0.071	0.133	0.068
		1	0.074	0.054	0.078	0.066	0.075	0.056
		3	0.056	0.054	0.069	0.067	0.060	0.057
	10	0.3	0.058	0.054	0.070	0.067	0.061	0.0578
		1	0.058	0.054	0.070	0.067	0.061	0.057
		3	0.058	0.054	0.069	0.067	0.061	0.057
20	1	0.3	0.119	0.068	0.108	0.054	0.117	0.063
		1	0.070	0.054	0.065	0.048	0.070	0.052
		3	0.054	0.052	0.052	0.049	0.054	0.052
	10	0.3	0.056	0.052	0.053	0.049	0.056	0.052
		1	0.056	0.052	0.053	0.049	0.056	0.052
		3	0.055	0.052	0.053	0.049	0.055	0.052
50	1	0.3	0.080	0.048	0.076	0.041	0.077	0.042
		1	0.051	0.041	0.047	0.036	0.047	0.036
		3	0.041	0.039	0.037	0.035	0.037	0.036
	10	0.3	0.042	0.039	0.038	0.035	0.038	0.036
		1	0.042	0.039	0.038	0.035	0.038	0.036
		3	0.041	0.039	0.038	0.035	0.038	0.036