An Introduction to Bayesian Empirical Likelihood

Farzana Jahan

December 7, 2017

Outline

Introduction

- Empirical Likelihood (EL) is a non-parametric analogue of data likelihood which does not require assumption regarding the underlying distribution of the data, but possesses the properties of parametric likelihood.
- Bayesian Empirical Likelihood (BEL) approach increases the flexibility of EL allowing the incorporation of prior knowledge with the data component.

Empirical Likelihood

- According to[1], EL is a technique for forming hypothesis tests and confidence regions based on nonparametric likelihood ratios
- EL provides likelihood inferences without assuming a parametric family

  For data $X_i \overset{\text{iid}}{\sim} F$

  $$L(F) = \prod_{i=1}^{n} (F(X_i) - F(X_i-)) = \prod_{i=1}^{n} p_i \textit{ is the nonparametric likelihood}$$

  $L(F)$ is the probability of getting exactly the observed sample values $x_1, x_2, ..., x_n$ from the CDF F.

[1] Art B Owen. "Empirical likelihood ratio confidence intervals for a single functional". In: *Biometrika* 75.2 (1988), pp. 237–249.

Non Parametric Likelihood Ratio

- In parametric likelihood methods, we construct confidence intervals of the form,
$$\left\{\theta | \frac{L(\theta)}{L(\theta_0)} \geqslant c\right\}$$
where the threshold c determines the confidence level.
- In parametric statistics, the parameters determine the distribution; in nonparametric statistics, we estimate the CDF directly using the empirical CDF, which is also the nonparametric maximum likelihood estimator of F

Profile Empirical Likelihood Ratio

- The analogous concept to a likelihood ratio is:

$$R(F) = \frac{L(F)}{L(F_0)}$$

  when, the data contains no ties and F places probability $p_i \geqslant 0$ on the value $X_i \in R$, then $L(F) = \prod_{i=1}^{n} p_i$ and $p_i = P_F(X = x_i)$. So,

$$R(F) = \prod_{i=1}^{n} np_i$$

- $R(F)$ is referred to as the Profile Empirical Likelihood Ratio

Profile Empirical Likelihood Ratio

- If $X_i = X_j$ for $i \neq j$, we say X_i and X_j are tied.
- If the data contains some ties, suppose the data contains k distinct values z_j arising $n_j \geqslant 1$ times in the sample and has probability p_j under F
- Then, the profile Empirical likelihood function can be defined as:

$$R(F) = \prod_{i=1}^{k} \left(\frac{np_j}{n_j} \right)^{n_j}$$

Profile Empirical Likelihood Ratio

- We define profile ELR in terms of observation specific weights $w_i \geqslant 0$ for $i = 1, 2, ..., n$. The weights are chosen so that $\sum_i w_i = p_j$ over all i with $X_i = z_j$.
- The EL of F in terms of weight becomes, $L(F) = \prod_{i=1}^{n} w_i$
- Then, the profile Empirical likelihood function can be defined as:

$$R(F) = \prod_{i=1}^{n} nw_i$$

where $w_i \geqslant 0, \sum_{i=1}^{n} w_i \leqslant 1$ and F puts probability $\sum_{X_i = X_j} w_j$ on X_i

Properties of EL

EL has a few properties of parametric likelihood:

- EL gives a data determined shape for confidence regions for two or more parameters of interest
- asymptotic distribution holds for the log likelihood ratio using EL[2]
- Bartlett's correction applies to EL[3]

[2] Samuel S Wilks. "The large-sample distribution of the likelihood ratio for testing composite hypotheses". In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62.

[3] Thomas DiCiccio, Peter Hall, and Joseph Romano. "Empirical likelihood is Bartlett-correctable". In: *the Annals of Statistics* (1991), pp. 1053–1061.

EL for a single functional

- EL can be constructed for different parameters of interest such as: mean, median, regression parameters etc.
- When, we are interested in inference concerning some functional of F, say, $\theta(F)$ and $\theta(F)$ can be determined by an estimating equation $f(y_i, \theta)$, then the empirical likelihood ratio function is defined by:

$$R(\theta) = max\left\{ \prod_{i=1}^{n} nw_i | \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i f(y_i, \theta) = 0 \right\}$$

- Estimating Equations
  $f(y_i, \theta) = Y - \theta$ for mean
  $f(y_i, \theta) = (Y - X^T\theta)X$ for regression and so on.

EL for Mean

- We will illustrate the ideas behind empirical likelihood by deriving the empirical likelihood for the mean
- Empirical Likelihood for mean:

$$R(\mu) = max\{\prod_{i=1}^{n} nw_i | \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i Y_i = \mu\}$$

- In order to compute $R(\theta)$ we can equivalently maximize $\sum_{i=1}^{n} logw_i$ subject to $\sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i Y_i = \mu$
- Note that because log is a concave function, a unique global maximum will exist

EL for Mean

- We may solve for the optimum values of $\{w_i\}$ using Lagrange multipliers; where our Lagrangian function G is

$$G = \sum_{i=1}^{n} logw_i + n\lambda \sum_{i=1}^{n} w_i(Y_i - \mu) - \gamma(\sum_{i=1}^{n} w_i - 1)$$

- Hence, $\frac{\delta G}{\delta w_i} = \frac{1}{w_i} - n\lambda(y_i - \mu) + \gamma$
  $0 = \sum_{i=1}^{n} w_i \frac{\delta G}{\delta w_i} = n - 0 + \gamma$
- Thus, $\gamma = -n$ and λ satisfies ,

$$\frac{1}{n} \frac{y_i - \mu}{1 + \lambda(y_i - \mu)} = 0$$

- There is no closed form solution for λ, so it should be solved by numerical search. In this case, λ will be solved by using Newton-Lagrange algorithm.

EL for Mean in R

- There is a package in R that solves empirical likelihood optimization problems called "emplik" developed by Mai Zhou in 2016
- The package details are available in: https://CRAN.R-project.org/package=emplik
- el.test from "emplik" package is used to calculate the -2 times log empirircal likelihood for any test value for mean and performs the required optimization.
- Demo in R

Bayesian Empirical Likelihood

- Likelihood being the central of Bayesian inference, is used to update prior beliefs and yields posterior inference.
- [4] argued that EL can be used in place of a density and, when multiplied by prior can yield the posterior distribution of interest in a Bayesian analysis.
- The author explored the characteristics of Bayesian inference using EL in spite of a parametric density.

[4] Nicole A Lazar. "Bayesian empirical likelihood". In: *Biometrika* 90.2 (2003), pp. 319–326.

Bayesian Empirical Likelihood

- To proceed with the Bayesian posterior, the profile EL function is chosen as the likelihood part of the Bayes theorem, and a prior on the functional θ, $p(\theta)$ is imposed
- The posterior of interest would be:

$$f(\theta|y_i) \propto p(\theta)EL(\theta)$$

- For $y_1, y_2, ..., y_n$ from some unknown F, let, θ be the parameter of interest for inference,$f(y_i, \theta)$ be the estimating equation, the empirical likelihood ratio function can be denoted by

$$W_E(\theta) = 2\sum_{i=1}^{n} log\{1 + \lambda^T f(y_i, \theta)\}$$

where the λ satisfies $\sum_{i=1}^{n} \frac{f(y_i,\theta)}{1+\lambda^T f(y_i,\theta)} = 0$

Bayesian Empirical Likelihood for mean

- For a specific case of one-dimensional parameter of interest, μ from a data set $y_, y_2, ..., y_n$, the profile empirical likelihood would be:

$$W_E(\mu) = 2\sum_{i=1}^{n} log\{1 + \lambda(y_i - \mu)\}$$

where λ satisfies: $\sum_{i=1}^{n} \frac{y_i - \mu}{1+\lambda(y_i - \mu)} = 0$

- The posterior of μ, $f(\mu|y_i) \propto EL(\mu)p(\mu)$
- Bayesian Empirical Likelihood method for calculating posterior is often referred to as "Semi-parametric method".
- Demo in R

Previous Research involving BEL

- [5] introduced Bayesian Exponentially Tilted Empirical Likelihood (BETEL) exploring the Bayesian analysis of moment condition models through EL.
- [6] applied BEL in quantile regression.
- [7] used EL in approximate Bayesian computation and proposed a new algorithm (Bayesian computation via empirical likelihood, BCel) by replacing the data likelihood by EL.

[5] Susanne M Schennach. "Bayesian exponentially tilted empirical likelihood". In: *Biometrika* 92.1 (2005), pp. 31–46.

[6] Yunwen Yang, Xuming He, et al. "Bayesian empirical likelihood for quantile regression". In: *The Annals of Statistics* 40.2 (2012), pp. 1102–1131.

[7] Kerrie L Mengersen, Pierre Pudlo, and Christian P Robert. "Bayesian computation via empirical likelihood". In: *Proceedings of the National Academy of Sciences* 110.4 (2013), pp. 1321–1326.

Previous Research involving BEL

- [8] applied BEL in construction of hierarchical spatial models for small area estimation.
- [9] developed a fully Bayesian framework for comparison and estimation in statistical models using BETEL.
- [10] developed an efficient Hamiltonian Monte Carlo method of sampling from BEL based posterior distribution of the paramers of interest

[8] Aaron T Porter, Scott H Holan, and Christopher K Wikle. "Bayesian semiparametric hierarchical empirical likelihood spatial models". In: *Journal of Statistical Planning and Inference* 165 (2015), pp. 78–90.

[9] Siddhartha Chib, Minchul Shin, and Anna Simoni. "Bayesian Estimation and Comparison of Moment Condition Models". In: *Journal of the American Statistical Association* just-accepted (2017).

[10] Sanjay Chaudhuri, Debashis Mondal, and Teng Yin. "Hamiltonian Monte Carlo sampling in Bayesian empirical likelihood computation". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79.1 (2017), pp. 293–320.

Future work plan

The proposed title of my PhD thesis is, "Bayesian Empirical Likelihood Methods for Big Data Analysis".
I will try to address the following research questions:

- How do existing BEL methods perform for big data analysis (high dimension and very large sample size)?
- How are BEL methods related other existing non parametric Bayesian methods for big data?
- Can BEL methods be improved to increase its performance in the context of Big data?
- How can BEL methods be developed for more complex hierarchical space-time models, mixed models and mixture models?

Thank You!