Empirical likelihood method in statistical inference 1

Hiroaki Ogata

at Université Libre de Bruxelles

March 9-12, 2010

Introduction

What is empirical likelihood?

- Empirical likelihood is a nonparametric method of inference based on a data-driven likelihood ratio function.
- It allows the data analyst to use likelihood methods without assuming that the data come from a known family of distributions.

Introduction

About likelihood method

- Likelihood method is known to be efficient. In other words, likelihood ratio tests have good power properties. Those tests can in turn be used to construct short confidence intervals or small confidence regions of the parameter.

- When we deal with the likelihood method, we usually consider parametric one.

Introduction

- In parametric likelihood methods $\cdots$

We suppose that the joint distribution of all available data has a known form.

Ex.

- Suppose that $X, X_1, \ldots, X_n \in \mathbb{R} \overset{i.i.d.}{\sim} Po(\lambda)$

$$\Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

- Suppose that $X, X_1, \ldots, X_n \in \mathbb{R} \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Introduction

Problem of parametric likelihood inference

- We might NOT know which parametric family to use.

$$\Downarrow$$

- Such misspecification can cause likelihood-based estimates to be inefficient.
- What may be worse is that the corresponding confidence intervals can fail completely!

Introduction

Nonparametric inference

- Nonparametric methods give confidence intervals and tests with validity NOT depending on strong distributional assumptions.

$\Downarrow$

- Empirical likelihood method combines
  - Reliability of the nonparametric methods
  - Flexibility and effectiveness of the likelihood approach !

Empirical cumulative distribution function

Definition

Let $X_1, \ldots, X_n \in \mathbb{R}$. The *empirical cumulative distribution function (ECDF)* of $X_1, \ldots, X_n$ is

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(-\infty, X_i]}(x)$$

for $-\infty < x < \infty$.

Here, $\mathbb{I}_A(x) = \begin{cases} 1 & (x \in A) \\ 0 & (x \notin A) \end{cases}$ is indicator function.

Nonparametric likelihood

Definition

Given $X_1, \ldots, X_n \in \mathbb{R}$, assumed independent with common cumulative distribution function (CDF) F_0. The *nonparametric likelihood* of the CDF F is

$$L(F) = \prod_{i=1}^{n} (F(X_i) - F(X_i-))$$

Here, $F(x) = \Pr(X \leq x)$ and $F(x-) = \Pr(X < x)$.

So, $\Pr(X = x) = F(x) - F(x-)$.

Properties of nonparametric likelihood

- If F is continuous distribution, then $L(F) = 0$.
- To have a positive nonparametric likelihood, a distribution F must place positive probability on every one of the observed values.

Question

What distribution maximizes the nonparametric likelihood?

Nonparametric maximum likelihood estimate

Theorem

Let $X_1, \ldots, X_n \in \mathbb{R}$ be independent random variables with a common CDF F_0. Let F_n be their ECDF and let F be any CDF. If $F \neq F_n$, then

$$L(F) < L(F_n).$$

- This theorem says ECDF F_n maximizes nonparametric likelihood $L(\cdot)$.

Parametric likelihood ratio

- In parametric inference we may base hypothesis tests and confidence regions on the likelihood ratio.

- Denote a parametric likelihood by $L^{(P)}(\eta)$ and maximum likelihood estimator (MLE) by $\hat{\eta}$, that is,

$$\hat{\eta} = \arg\max_{\eta} L^{(P)}(\eta)$$

- For a certain η_0, if $L^{(P)}(\eta_0)$ is much smaller than $L^{(P)}(\hat{\eta})$, then we reject the hypothesis that $\eta = \eta_0$, and exclude η_0 from our confidence region for η.

Parametric likelihood ratio

- Wilks's theorem

$$-2\log\left(\frac{L^{(P)}(\eta_0)}{L^{(P)}(\hat{\eta})}\right) \xrightarrow{d} \chi_p^2 \quad \text{under} \quad \eta = \eta_0$$

where p is the dimension of the parameter η_0.

- We can construct the asymptotic $100(1-\alpha)\%$ confidence region as

$$\left\{\eta \;\middle|\; -2\log\left(\frac{L^{(P)}(\eta)}{L^{(P)}(\hat{\eta})}\right) < \chi_p^{2,1-\alpha}\right\}$$

where $\chi_p^{2,1-\alpha}$ is $(1-\alpha)$-quantile of χ_p^2.

Nonparametric likelihood ratio

- Similar to the parametric case, we may also use the ratio of the nonparametric likelihood as a basis for hypothesis test and confidence intervals.
- For a distribution F, define the *nonparametric likelihood ratio* as

$$R(F) = \frac{L(F)}{L(F_n)}.$$

- Suppose that we are interested in a parameter $\boldsymbol{\theta} = T(F)$ for some function T of distributions.

$$T : \mathcal{F} \to \mathbb{R}^p \qquad (\mathcal{F} \text{ is a set of distributions})$$

Example of θ

- Mean

$$\theta = T(F) = \int x dF(x)$$

- Variance

$$\theta = T(F) = \int x^2 dF(x) - \left(\int x dF(x) \right)^2$$

- α-quantile

$$\theta = T(F) = F^{-1}(\alpha)$$

Profile likelihood ratio function

- We define the profile likelihood ratio function of $\boldsymbol{\theta}$ as

$$\mathcal{R}(\boldsymbol{\theta}) = \sup\{R(F) \mid T(F) = \boldsymbol{\theta}, F \in \mathcal{F}\}.$$

- Empirical likelihood hypothesis tests reject

$$H_0 : T(F_0) = \boldsymbol{\theta}_0 \quad \text{when} \quad \mathcal{R}(\boldsymbol{\theta}_0) < r_0$$

  where r_0 is certain threshold value.

- Empirical likelihood confidence regions are of the form

$$\{\boldsymbol{\theta} \mid \mathcal{R}(\boldsymbol{\theta}) \geq r_0\}$$

EL for univariate mean

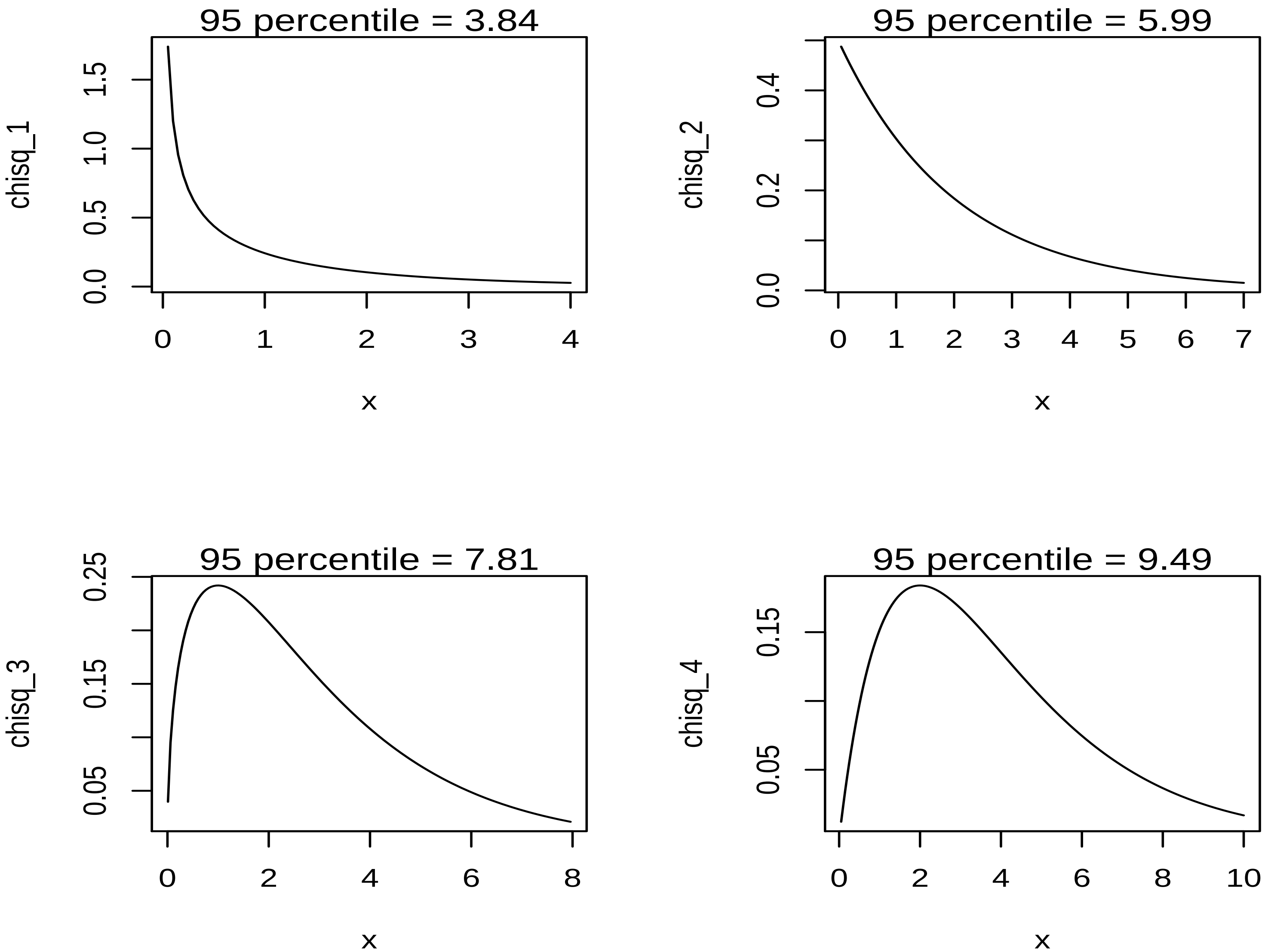
Theorem (Univariate ELT)

Let $X_1, \ldots, X_n \in \mathbb{R}$ be independent random variables with common distribution F_0. Let $\mu_0 = \mathrm{E}(X_i)$, and suppose that $0 < \mathrm{Var}(X_i) < \infty$. Then

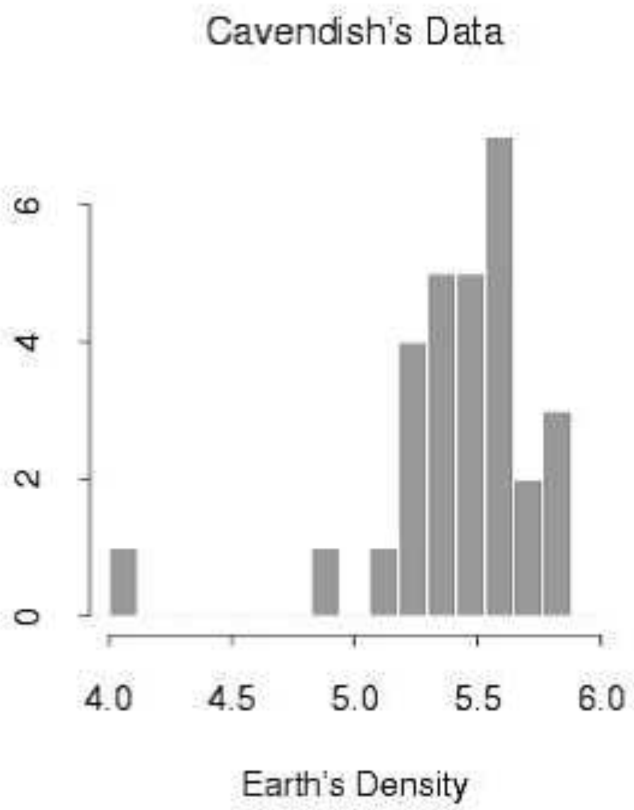
$$-2\log(\mathcal{R}(\mu_0)) \xrightarrow{d} \chi^2_{(1)} \qquad (n \to \infty).$$

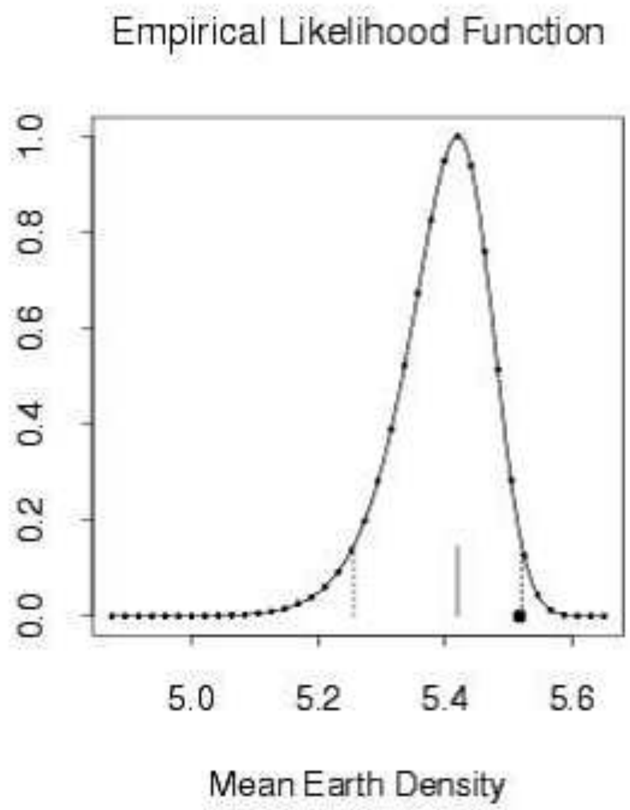
- This asymptotic distribution (chi-square) is same as the parametric one !
- We reject the value μ_0 at the α level when
$$-2\log(\mathcal{R}(\mu_0)) > \chi^{2,1-\alpha}_{(1)}.$$
- We construct $100(1-\alpha)\%$ confidence region as
$$\{\mu \mid -2\log(\mathcal{R}(\mu)) < \chi^{2,1-\alpha}_{(1)}\}.$$

Density functions of chi-square



Earth density estimation





- The left is the histgram of 29 determinations of the mean density of the earth. The right is the ELR function for the mean of these data, with the presently accepted value 5.517 marked. The 95% EL confidence interval extends from 5.256 to 5.521. (`http://www-stat.stanford.edu/~owen/empirical/`)

Computation of EL for a univariate mean

- If the distribution F places probability p_i on the sample $X_i \in \mathbb{R}$, then $\sum_{i=1}^n p_i = 1$ and $L(F) = \prod_{i=1}^n p_i$, so

$$R(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^{n} np_i.$$

- Then, EL ratio for a mean $\mu = \int x dF(x)$ is

$$\mathcal{R}(\mu) = \max_{\boldsymbol{p}} \left\{ \prod_{i=1}^{n} np_i \;\middle|\; \sum_{i=1}^{n} p_i X_i = \mu, p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}$$

where $\boldsymbol{p} = (p_1, \ldots, p_n)'$.

Trivial case

Let the ordered sample values be $X_{(1)} \leq \ldots \leq X_{(n)}$.

- The case of $\mu < X_{(1)}$ or $\mu > X_{(n)}$
  There are no weights p_i's which satisfy

$$\sum_{i=1}^{n} p_i X_i = \mu, \; p_i \geq 0, \; \sum_{i=1}^{n} p_i = 1.$$

  We take $\mathcal{R}(\mu) = 0$ by convention in this case.

- Similarly, we take

$$\mathcal{R}(\mu) = \begin{cases} 0 & (\mu = X_{(1)} < X_{(n)}, \; \mu = X_{(n)} > X_{(1)}) \\ 1 & (\mu = X_{(1)} = X_{(n)}) \end{cases}$$

Nontrivial case

Let us consider the nontrivial case $X_{(1)} < \mu < X_{(n)}$.

- We maximize $\prod_i np_i$, or equivalently

$$\sum_{i=1}^{n} \log(np_i)$$

over $p_i \geq 0$ subject to the constraints that

$$\sum_{i=1}^{n} p_i(X_i - \mu) = 0 \quad \text{and} \quad \sum_{i=1}^{n} p_i - 1 = 0.$$

Method of Lagrange multipliers

Method of Lagrange multipliers is used when we consider the maximization problem under some constraints.

- Write

$$G = \sum_{i=1}^{n} \log(np_i) - n\lambda \sum_{i=1}^{n} p_i(X_i - \mu) + \gamma \left(\sum_{i=1}^{n} p_i - 1 \right)$$

  where λ and γ are Lagrange multipliers.

- Set

$$0 = \frac{\partial G}{\partial p_i} = \frac{1}{p_i} - n\lambda(X_i - \mu) + \gamma$$

  for $i = 1, \ldots, n$.

Weights p_i's

- So

$$0 = \sum_{i=1}^{n} p_i \frac{\partial G}{\partial p_i} = n + \gamma$$

from which $\gamma = -n$.

- We may therefore write

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda(X_i - \mu)}.$$

for $i = 1, \ldots, n$. The value of λ may be found by numerical search

Lagrange multiplier λ

- Lagrange multiplier $\lambda = \lambda(\mu)$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\frac{X_i - \mu}{1+\lambda(X_i - \mu)} = 0. \qquad (*)$$

- The left side of $(*)$ is strictly decreasing in λ (Differentiate w.r.t. λ).

- Therefore, we can use a bisection approach to obtain the solution of (1). We may start the search from

$$\frac{1-n^{-1}}{\mu - X_{(n)}} < \lambda(\mu) < \frac{1-n^{-1}}{\mu - X_{(1)}}$$