Empirical Likelihood

Patrick Breheny

September 20

Introduction

- We will discuss one final approach to constructing confidence intervals for statistical functionals
- The idea is to extend the tools of maximum likelihood directly to the nonparametric case
- In parametric likelihood methods, we construct confidence intervals of the form

$$\left\{ \theta \,\middle|\, \frac{L(\theta)}{L(\hat{\theta})} \geq c \right\},$$

where the threshold c determines the confidence level

The nonparametric likelihood ratio

- 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
- The analogous concept to a likelihood ratio is

$$R(F) = \frac{L(F)}{L(\hat{F})},$$

where $L(F) = \prod_i w_i$ and $w_i = \mathbb{P}_F(X = x_i)$

Empirical likelihood confidence regions

- How does this help us to find a confidence interval for $\theta = T(F)$?
- Define

$$\mathcal{R}(\theta) = \sup_F \{R(F) | T(F) = \theta, F \in \mathcal{F}\}$$

- In the terminology of likelihood theory, this is a *profile likelihood* ratio, where the numerator is the likelihood of the parameter of interest, maximized over the nuisance parameters
- Empirical likelihood confidence regions are then of the form

$$\{\theta | \mathcal{R}(\theta) > r\}$$

Empirical likelihood of the mean

- We will illustrate the ideas behind empirical likelihood by deriving the empirical likelihood confidence interval for the mean
- We immediately encounter a challenge: if we let $\mathcal{F}$ be the set of all possible distributions, our confidence interval is infinitely wide
- In order to eliminate this problem, we must restrict $\mathcal{F}$ in some way
- Before we move on, however, note that we may not need to restrict $\mathcal{F}$ if dealing with a more robust statistic such as the median

Restricting $\mathcal{F}$ to the sample

- One natural approach is to restrict the support of $\mathcal{F}$ to include only the points $\{x_i\}$
- Once this restriction is put in place, calculation of $\mathcal{R}(\theta)$ amounts to maximizing the nonparametric likelihood over a finite number (n) of weights $\{w_i\}$, subject to certain restrictions and constraints
- We will discuss the details of this a little later

Asymptotic distribution of $\mathcal{R}(\theta)$

- It turns out that, asymptotically, $\mathcal{R}(\theta)$ behaves very similarly to the parametric likelihood ratio
- **Theorem:** Let $X \overset{\text{iid}}{\sim} F_0$ and $\theta_0 = \mathbb{E}(X)$, and suppose $\mathbb{V}(X) \in (0, \infty)$. Then

$$-2 \log \mathcal{R}(\theta_0) \xrightarrow{\text{d}} \chi_1^2$$

- The above holds for any sufficiently smooth (*i.e.* differentiable) statistical functional, given appropriate regularity conditions

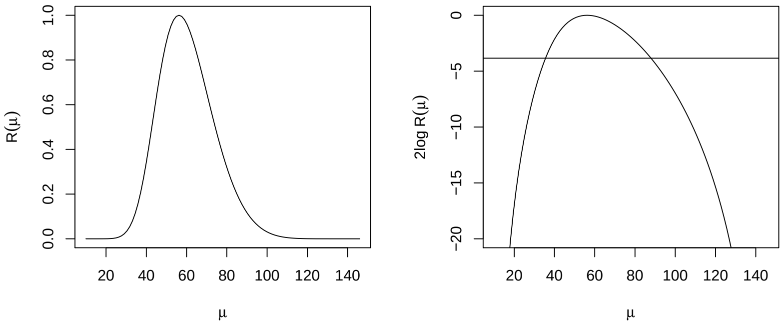
EL hypothesis tests and confidence intervals

- The preceding theorem allows us to construct hypothesis tests by calculating the area under the χ_1^2 curve outside $-2\log\mathcal{R}(\theta_0)$
- It also allows for the construction of confidence intervals of the form:

$$\{\theta| -2\log\mathcal{R}(\theta) < \chi^2_{1,1-\alpha}\}$$

Example: Rat survival

Let's apply empirical likelihood to our study of survival in rats that was introduced in the previous lecture:



Comparison

Comparing this interval to our intervals from the last lecture:

Normal	28.5	84.0
t	23.6	88.9
Bootstrap-t	33.4	125.1
Percentile	32.1	87.7
BC_a	37.3	90.9
EL	35.6	88.0

Setup

- In order to compute $\mathcal{R}(\theta)$, we have to maximize $\prod_i nw_i$ – or equivalently, $\sum_i \log(nw_i)$ – over $\{w_i\}$ subject to the following constraints:

$$w_i > 0 \quad \forall\, i$$
$$\sum_i w_i = 1$$
$$\sum_i w_i x_i = \theta$$

- Note that because $\log$ is a strictly concave function, a unique global maximum will exist

Lagrange multipliers

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

$$G = \sum_i \log(nw_i) - n\lambda \sum_i w_i(x_i - \theta) - \gamma \left(\sum_i w_i - 1 \right)$$

- Thus, $\gamma = n$ and λ satisfies

$$\frac{1}{n} \sum_i \frac{x_i - \theta}{1 + \lambda(x_i - \theta)} = 0$$

Solving for λ

- There is no closed form solution for λ, so we must use some form of univariate root-finding algorithm such as Brent's method (used by `uniroot`)
- To begin the search, we need an initial bracket for λ
- For the mean, this can be obtained by setting the weight of the largest observation to 1, and then setting the weight of the smallest observation to 1:

$$\lambda \in \left(\frac{1-n}{n(x_{(n)}-\theta)}, \frac{1-n}{n(x_{(1)}-\theta)} \right)$$

Determination of the confidence interval

- Once we have λ, the weights follow from

$$w_i = \frac{1}{n} \frac{1}{1 + \lambda(x_i - \theta)}$$

  and the likelihood can be calculated.

- Solving for the endpoints (θ_L, θ_U) of a confidence interval can either be interpolated from the calculation of $\mathcal{R}(\theta)$ or solved via a similar sort of Lagrangian technique

Homework

- **Homework:** Show that, for $\theta = \mathbb{E}(X)$,

$$-2\log\mathcal{R}(\theta_0) \xrightarrow{\mathsf{d}} \chi^2_1$$

- Hint: do this in two parts. For the first part, take a Taylor series expansion of

$$\frac{1}{n}\sum_i \frac{x_i - \theta}{1 + \lambda(x_i - \theta)} = 0$$

  about $\lambda = 0$ to show that $\lambda \approx (\bar{x} - \theta)/S$, where $S = n^{-1}\sum_i (x_i - \theta)^2$

- In the second part, use the above approximation to show that $-2\log\mathcal{R}(\theta_0) \approx n(\bar{x} - \theta_0)^2/S$ (Hint: use the fact that when $x \approx 0$, $\log(1 + x) \approx x - \frac{1}{2}x^2$)