Adv Stat Methods II Tutorial 1

Vania Lin

2023-05-30

I()

- In function formula, it is used to inhibit the interpretation of operators such as "+", "-", "*" and "^" as formula operators, so they are used as arithmetical operators.
- E.g.

$(x+y)^2=x^2+2xy+y^2$ (mathematical/arithmetical)

$=x+y+x:y$ (symbolic)

Background

- For public health surveillance purposes, a study has been carried out to assess the prevalence of obesity (defined as BMI>25) among male HKU undergraduates. The study took a random sample of 185 undergraduate males in May 2023, and assessed each of their heights and weights. In total, 13 students were found to be obese.

- The log-likelihood for the prevalence θ is given by $\log l(\theta) = 13 log\theta + 172 \log(1-\theta)$. The MLE estimates for θ is 13/185 = 0.07. By using a normal approximation, the 95% confidence interval is (0.03, 0.11).

$$log\, l(\theta) = 13\, log(\theta) + 172\, log(1-\theta)$$

Questions

(a) Using the likelihood ratio method, obtain a 95% confidence interval for $\boldsymbol{\theta}$.

$$2\left|log\, l(\hat{\theta}) - log\, l(\theta)\right| \sim \chi_1^2$$

Questions

(a) Using the likelihood ratio method, obtain a 95% confidence interval for $\boldsymbol{\theta}$.

$$2\left|log\, l(\hat{\theta}) - log\, l(\theta)\right| \sim \chi_1^2$$

A method to derive confidence intervals based on likelihood ratios (Wilks' theorem) using the (asymptotic) property.
The 95th percentile of the χ_1^2 distribution is 3.84.
The values of θ for which $|\log l(\hat{\theta}) - \log l(\theta)| < 1.92$ will form a 95% confidence interval for θ. (slide 39, lecture 1)

Proof of Wilks' Theorem on LRT

This handout is intended to supply full details and re-cap of carefully defined notations of the argument given in class proving the asymptotic distributional convergence of the likelihood ratio test statistic of general null hypotheses restricting the values of a subset of parameter components to a chi-square with degrees of freedom equal to the number of components restricted. Throughout, we assume that the data $\mathbf{X} = \{X_i\}_{i=1}^n$ constitute an *iid* sample (of values in some Euclidean data-space) from a density $f(x, \vartheta)$ known except for unknown parameter $\vartheta \subset \Theta \subset \mathbf{R}^k$. We assume that the density f satisfies all of the regularity conditions previously needed to ensure that maximum likelihood estimators are locally unique, consistent, and asymptotically normal. These conditions include the restriction that Θ contain an open neighborhood of the true value ϑ_0 governing the data, and that Θ lies in some sufficiently small neighborhood of ϑ_0 not depending upon n. The likelihood for the data $\mathbf{X}$ is denoted by $L(\mathbf{X}, \vartheta)$ and the unrestricted Maximum Likelihood Estimator (MLE) for ϑ is $\hat{\vartheta}$.

Now consider the null hypothesis $H_0 : \vartheta_{0,j} = 0$ for $1 \le j \le r$, where $0 < r < k$ is fixed. Define the *restricted MLE* $\hat{\vartheta}^{res}$ as the maximizer of $L(\mathbf{X}, \vartheta)$ over parameter vectors $\vartheta \subset \Theta$ such that $\vartheta_{0,j} = 0$ for $1 \le j \le r$. We require a detailed set of notations designed to partition parameters, estimators, gradients, score statistics, and information matrices into parts respectively reflecting the first r and the last $k - r$ components. Under the null hypothesis, the parameter vector ϑ_0 and restricted MLE have the form

$$\vartheta_0 = \begin{pmatrix} \mathbf{0} \\ \vartheta_* \end{pmatrix} \quad , \quad \hat{\vartheta}^{res} = \begin{pmatrix} \mathbf{0} \\ \hat{\vartheta}_* \end{pmatrix} \quad , \quad \vartheta_{0*},\ \hat{\vartheta}_* \in \mathbf{R}^{k-r}$$

Next, denote by ∇_A, ∇_C respectively the gradient operator with respect to the first r and last $k - r$ component of ϑ. It is clear that the MLE definitions are equivalent to

$$\nabla \log L(\mathbf{X}, \hat{\vartheta}) \equiv \begin{pmatrix} \nabla_A \\ \nabla_C \end{pmatrix} \log L(\mathbf{X}, \hat{\vartheta}) = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \quad , \quad \nabla_C \log L(\mathbf{X}, \hat{\vartheta}^{res}) = \mathbf{0}$$

http://www.math.umd.edu/~slud/s701.S14/WilksThm.pdf

Questions

(a) Using the likelihood ratio method, obtain a 95% confidence interval for $\boldsymbol{\theta}$.

$$2\left|\log l(\hat{\theta}) - \log l(\theta)\right| \sim \chi_1^2$$

- $\hat{\theta}$=MLE=$\frac{13}{185}$ (observed; already known)
- $\left|\log l(\frac{13}{185}) - \log l(\theta)\right| < 1.92$

Questions

(b) Using the bootstrap method, obtain a 95% confidence interval for $\boldsymbol{\theta}$.

Questions

(b) Using the bootstrap method, obtain a 95% confidence interval for $\boldsymbol{\theta}$.

Bootstrap package ('boot') in R

boot(*data*, *statistic*, *R*) [generate bootstrap samples]

- *statistic* is a function returning the statistic of interest
  - there should be at least two arguments for the function, first: original data; second: indices which specify bootstrap samples
- *R* is the number of bootstrap replicate

boot.ci(*boot.out*, *conf = 0.95*, *type = "all"*) [construct bootstrap CI]

- *boot.out* is the object of class "boot" with bootstrap calculation
- *conf* is the confidence level for the confidence interval
- *type* selects the type of bootstrap confidence intervals, such as "*norm*", "*stud*", "*perc*", "*bca*"

Questions

(c) Suppose the study was also carried out in 7 other tertiary institutions. Their results are summarized below:

Institutions	#1	#2	#3	#4	#5	#6	#7
No. obese	18	21	10	11	10	17	12
No. male undergraduate	161	272	154	85	101	221	150

Estimate the overall prevalence using maximum likelihood method.

Questions

(c) Estimate the overall prevalence using maximum likelihood method.

- The log likelihood for θ is

$$\sum_{i=1}^{8} [x_i log\theta + (n_i - x_i) \log(1-\theta)]$$

Questions

(d) Suppose it was hypothesized that institutions which were able to recruit more participants (e.g. n > 200) may have a different prevalence of obesity. Estimate the relative difference using the maximum likelihood method.

You may assume that the obesity prevalence is $\boldsymbol{\theta}$ for schools with fewer participants, and $\boldsymbol{k\theta}$ for schools with more participants.

Questions

(d) Suppose it was hypothesized that institutions which were able to recruit more participants (e.g. n > 200) may have a different prevalence of obesity. Estimate the relative difference using the maximum likelihood method.

- The log likelihood for *k* and θ (<u>two unknown parameters</u>) is

$$\sum_{low}[x_i log\theta + (n_i - x_i)\log(1-\theta)] + \sum_{high}[x_i logk\theta + (n_i - x_i)\log(1-k\theta)],$$

or

$$\sum_{i=1}^{8}[x_i logk^{high}\theta + (n_i - x_i)\log\left(1 - k^{high}\theta\right)]$$

optim()

- By default optim performs minimization, but it will maximize if control$fnscale is negative.

- Methods:

- "Brent”: for one-dimensional problems only (i.e. for single value).

- "L-BFGS-B”: allows box constraints and each variable can be given a lower and/or upper bound (i.e. can be used to deal with vectors). The initial value must satisfy the constraints.

Questions

(e) When the sample size is large, according to maximum likelihood theory

$$\hat{\theta} \sim N(\theta, I^{-1}(\theta)),$$

where $I^{-1}(\theta)$ is the information matrix

$$I(\theta) = -E[\frac{\partial^2 logL(\theta)}{\partial\theta\partial\theta'}]$$

$\frac{\partial^2 logL(\theta)}{\partial\theta\partial\theta'}$ is the second derivative of the log-likelihood, also named Hessian, which can be obtained by setting "hessian=T" in the optim function in R.

Compute the standard error for the estimated prevalence of obesity <u>in the first tertiary institution</u> and calculate its 95% confidence interval.

Questions

(e) Compute the standard error for the estimated prevalence of obesity in the first tertiary institution and calculate its 95% confidence interval

$$\widehat{\boldsymbol{\theta}} \sim \boldsymbol{N}(\boldsymbol{\theta}, \boldsymbol{I}^{-1}(\boldsymbol{\theta}))$$

- $\boldsymbol{I}^{-1}(\boldsymbol{\theta})$: information matrix (as the variance), which is the negative inverse of the Hessian matrix
- SE could be obtained by taking squared root of the variance.

Questions

(f) Referring to (d), compute the 95% confidence interval for k and test the hypothesis H0: k=1.

[Hint: use solve() to compute the inverse of a matrix]

Questions

(f) Referring to (d), compute the 95% confidence interval for k and test the hypothesis H0: k=1.

[Hint: use solve() to compute the inverse of a matrix]

Mathematical details [edit]

Under the Wald test, the estimated $\hat{\theta}$ that was found as the maximizing argument of the unconstrained likelihood function is compared with a hypothesized value θ_0. In particular, the squared difference $\hat{\theta} - \theta_0$ is weighted by the curvature of the log-likelihood function.

Test on a single parameter [edit]

If the hypothesis involves only a single parameter restriction, then the Wald statistic takes the following form:

$$W = \frac{(\hat{\theta} - \theta_0)^2}{\operatorname{var}(\hat{\theta})}$$

which under the null hypothesis follows an asymptotic χ^2-distribution with one degree of freedom. The square root of the single-restriction Wald statistic can be understood as a (pseudo) *t*-ratio that is, however, not actually *t*-distributed except for the special case of linear regression with normally distributed errors.[12] In general, it follows an asymptotic *z* distribution.[13]

$$\sqrt{W} = \frac{\hat{\theta} - \theta_0}{\operatorname{se}(\hat{\theta})}$$

where $\operatorname{se}(\hat{\theta})$ is the standard error of the maximum likelihood estimate (MLE), the square root of the variance. There are several ways to consistently estimate the variance matrix which in finite samples leads to alternative estimates of standard errors and associated test statistics and *p*-values.[14]

Questions

(g) Perform a likelihood ratio test for (f).

- Compare the loglikelihood with/without the parameter *k*.