Imperial College London

Problem Sheet 3

MATH50011
Statistical Modelling 1

Week 4

Lecture 7 (Proof of MLE Consistency and Asymptotic Normality)

1. **(Challenge)** In the lecture notes, we saw that MLEs are asymptotically normal and sketched a proof of this (subject to regularity conditions). Many other estimators are also the solutions to estimating equations.

   Let $X_1, \dots, X_n$ be i.i.d. real-valued random variables and suppose that we wish to estimate the value of $\theta_0 \in \mathbb{R}$ defined as the unique $E[\psi(X_1, \theta)] = 0$ for a twice continuously differentiable function $\psi : \mathbb{R}^2 \to \mathbb{R}$. Define $\hat{\theta}_n$ as the unique solution to $\sum_{i=1}^{n} \psi(X_i, \theta) = 0$.

   Sketch a proof that $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N(0, \sigma^2(\theta_0))$ and find an expression for $\sigma^2(\theta_0)$. At what steps in your proof sketch do you need additional assumptions required to justify an operation?

2. **(Challenge)** In the notation of the `R` lab question (see below), define the one-step estimator
$$\hat{\theta}_n^{(1)} = T_n + I_n(T_n)^{-1} U_n(T_n).$$
   Suppose that T_n is asymptotically normal and that $I_n(T_n)$ is consistent for the Fisher information $I_f(\theta)$. Show that
$$\sqrt{n}(\hat{\theta}_n^{(1)} - \theta_0) \to_d N(0, I_f(\theta_0)^{-1}).$$
   Hint: use a first-order Taylor expansion of $U_n(\theta)$.

Lecture 8 (Confidence Intervals)

3. Dr. Jetson asked a random sample of 10000 UK households whether or not they own a robotic vacuum cleaner. She finds that 1300 of the households own a robotic vacuum and the other 8700 do not. Based on this data, she estimates that 13% of UK households own a robotic vacuum with a 95% confidence interval of 12.3% to 13.7%. Dr. Jetson tells you that

   > "There is a 95% probability that between 12.3% and 13.7% of UK households own a robotic vacuum cleaner."

   What is the main problem with the above statement? Provide a correct description of the confidence interval suitable for a non-statistician.

4. A random sample of 11 components in a factory is collected. The length in cm of each component is recorded below

   3.26 1.76 1.63 1.79 2.43 0.88 0.99 1.12 4.56 2.11 2.73

   Assume that the lengths are normally distributed with mean μ and variance σ^2. Construct a 99% confidence interval for μ.

5. Let $Y_1, \dots, Y_n$ be i.i.d. Exp(λ), where $\lambda > 0$ is unknown.

   (a) Show that $2\lambda \sum_{i=1}^{n} Y_i$ has a χ^2-distribution with $2n$ degrees of freedom;

   (b) Derive a $(1 - \alpha) \times 100\%$ confidence interval for λ;

   (c) Using the following observations, compute a 95% confidence interval for λ.

   1.04 1.39 0.1 2.04 4.73 0.89 0.51 0.89 0.66 0.93

   (Note: for $X \sim \chi^2_{20}$, $P(X \leq 9.59) = 0.025$ and $P(X \leq 34.17) = 0.975$.)

6. **(Challenge)** Find an approximate 95% confidence interval for the odds that a randomly selected UK household owns a robotic vacuum based on the data in Question 3. (Hint: use the delta method.)

7. Use the Bonferroni correction to find a 95% confidence region for (μ, σ^2) based on a random sample $X_1, \dots, X_n$ from a $N(\mu, \sigma^2)$ distribution. Apply your result to construct a 95% confidence region for (μ, σ^2) based on the data in Question 4.

R lab: One-Step Estimators

This exercise is intended to reinforce concepts through use of the R software package.

In the notes, we saw that numerical methods can facilitate maximisation of the (log) likelihood. In this lab, we illustrate how a simple one-step update to an initial estimator can lead to an accurate approximation of the MLE. The step we take is based on Newton's method.

Suppose that $X_1, \dots, X_n$ are iid with pdf $f_\theta(x)$. Define

$$U_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f_\theta(X_i)$$
$$I_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2}{\partial\theta^2} \log f_\theta(X_i)$$

The one-step estimator is defined as $\hat\theta_n^{(1)} = T_n - I_n(T_n)^{-1}U_n(T_n)$, where T_n is an initial estimator of θ. If T_n is an asymptotically normal estimator of θ, then

$$\sqrt{n}(\hat\theta_n^{(1)} - \theta) \to_d N(0, I_f(\theta)^{-1}).$$

You will prove this in the next problem sheet.

8. In this exercise, you will implement a simulation study to explore the behavior of the one-step estimator for the location parameter θ of the Cauchy(θ) distribution with pdf

   $$f_\theta(x) = \frac{1}{\pi\,[1 + (x-\theta)]^2} \quad -\infty < x < \infty, -\infty < \theta < \infty.$$

   Note that $f_\theta(x)$ is symmetric about θ. However, $E_\theta(X)$ does not exist for the Cauchy distribution so the sample mean would be an awful estimator here. Instead, we will use the sample median as an initial estimator of θ.

   After drawing $X_1, \dots, X_n$ i.i.d. Cauchy(θ), the sample median $\hat m_n$ will be computed and stored as an initial estimator. The values of $U_n(\hat m_n)$ and $I_n(\hat m_n)$ are then computed and used to construct a one-step estimator $\hat\theta_n^{(1)}$ based on $\hat m_n$. This experiment will be independently replicated a total of 1000 times, so that we can approximate the sampling distributions of $\hat m_n$ and $\hat\theta_n^{(1)}$.

The R code below implements the simulation study for $n = 10$ and $\theta = 0$.

```
set.seed(50011)
result.m <- logical(length = 1000)
result.t1 <- logical(length = 1000)
n <- 10
theta <- 0
for(i in 1:1000){
X <- rcauchy(n, location = 0)
m <- median(X)
U <- NULL
I <- NULL
t1 <- m - U/I
result.m[i] <- sqrt(n)*(m-theta)
result.t1[i] <- sqrt(n)*(t1-theta)
}
hist(result.m, freq=FALSE)
hist(result.t1, freq=FALSE)
```

Note that the command `set.seed(50011)` ensures that you obtain the same results each time you run this set of commands.

Type the above commands into an `R` script and then:

(a) Derive expressions for $U_n(\hat{m}_n)$ and $I_n(\hat{m}_n)$ in terms of `X` and `m`. Use your expressions to replace the appropriate `NULL` definitions in the for loop.

(b) Comment on why it is reasonable to store the values of $\sqrt{n}(\hat{m}_n - \theta)$ and $\sqrt{n}(\hat{\theta}_n^{(1)} - \theta)$ instead of $\hat{\theta}_n^{(1)}$ and $\hat{m}_n$.

(c) Explore how each histogram changes by increasing the value of `n` in this code to, e.g. $n = 30, 50, 100, 200, 500, 1000$. You might also compare other, say numerical, summaries (e.g. mean, variance, quantiles).

(d) Referring to your results from (c), comment on whether you prefer the sample median or one-step estimator for estimating θ in this setting.

Challenge Do your simulations provide evidence that $\sqrt{n}(\hat{\theta}_n^{(1)} - \theta)$ convergences in distribution to a $N(0, I_f(\theta)^{-1})$ random variable? Explain your answer using appropriate graphical and/or numerical evidence.