Imperial College London

Problem Sheet 3

MATH50011
Statistical Modelling 1

Week 3

Lecture 5 (Asymptotic Normality)

1. Prove that if $X_1, X_2, \ldots$ converges in probability to X and h is a continuous function, then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.

2. Suppose that $X_1, \ldots, X_n$ are iid with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Define $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ and $S_n^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

   (a) Show that S_n^2 is a consistent estimator of σ^2. Assume that all required higher order moments of X_i exist and are finite.

   (b) Use the result in (a) to show that

$$T_n = \sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \right) \to_d N(0,1)\,.$$

3. Suppose that $X_1, \ldots, X_n$ are iid strictly positive random variables with $E(\log X_i) = \mu$ and $\text{Var}(\log X_i) = \sigma^2$. Use the delta method to derive the asymptotic normality of the geometric mean $G_n = \left(\prod_{i=1}^n X_i\right)^{1/n}$.

4. Suppose that $X_1, \ldots, X_n$ are iid Uniform$(0, \theta)$ and define $T_n = \max(X_1, \ldots, X_n)$. Find a sequence $a_n = n^k$ for some k such that $a_n(T_n - \theta) \to_d Z$. What is the distribution of Z?

5. Does $\sqrt{n}(T_n - \theta) \to_d N(0, \sigma^2)$ imply that T_n is consistent for θ? If yes, prove this. Otherwise, provide a counterexample.

Lecture 6 (Maximum Likelihood)

6. Find the MLE for estimating θ based on a random sample $X_1, \dots, X_n$ from the following distributions
   (a) Bernoulli(θ); (see Example 8)
   (b) Poisson(θ);
   (c) Exponential(θ);
7. For the distributions in 6(a-c), find Z such that $\sqrt{n}(\hat{\theta}_n - \theta) \to_d Z$.
8. For the distributions in 6(a) and 5(b), find the MLE $\hat{\nu}_n$ of $\nu = g(\theta) = P_\theta(X_1 = 0)$ and show that $\sqrt{n}(\hat{\nu}_n - \nu) \to_d Z$. Find the distribution of Z in each case.
9. Suppose that we wish to estimate θ based on a random sample $X_1, \dots, X_n$ of Bernoulli(θ) random variables. However, we are only able to obtain a random sample $(Y_i, R_i), \dots, (Y_n, R_n)$ where the R_i's are iid Bernoulli(p_0) for known p_0 and $Y_i = R_i X_i$ for $i = 1, \dots, n$. Derive the MLEs $\hat{\theta}_a, \hat{\theta}_b$ and $\hat{\theta}_c$ for θ based on
   (a) The full data distribution of the X_i's;
   (b) The marginal distribution of the Y_i's;
   (c) The joint distribution of the (Y_i, R_i)'s.
10. Let T_n and U_n be estimators of θ such that

$$\sqrt{n}(T_n - \theta) \to_d N(0, \sigma_T^2)$$
$$\sqrt{n}(U_n - \theta) \to_d N(0, \sigma_U^2)\,.$$

The *asymptotic relative efficiency* of T_n with respect to U_n is σ_T^2/σ_U^2.

Find the asymptotic distributions of the MLEs in 9(b) and 9(c) and calculate the asymptotic relative efficiency of $\hat{\theta}_b$ to $\hat{\theta}_c$. Which of the MLEs do you prefer for estimating θ? Quantify the loss in efficiency of your preferred estimator to $\hat{\theta}_a$ that is based on the (unobserved) X_i's. Explain.

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.

11. 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.