Imperial College London

Problem Sheet 2 Solutions

MATH50011
Statistical Modelling 1

Week 3

Lecture 5 (Asymptotic Normality)

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

Solution. Let $\epsilon > 0$ be given. We want to show that

$$\lim_{n\to\infty} P(|h(X_n) - h(X)| < \epsilon) = 1.$$

By continuity of h, we know that there exists $\delta \equiv \delta(\epsilon)$ such that

$$|X_n - X| < \delta \Rightarrow |h(X_n) - h(X)| < \epsilon.$$

This implies that

$$P(|X_n - X| < \delta) \leq P(|h(X_n) - h(X)| < \epsilon).$$

We know that the right hand side is bounded above by one. Taking limits we find

$$1 = \lim_{n\to\infty} P(|X_n - X| < \delta) \leq \lim_{n\to\infty} P(|h(X_n) - h(X)| < \epsilon) \leq 1,$$

where we have used $X_n \to_p X$. Hence, we conclude that $h(X_n) \to_p h(X)$.

2. Suppose that $X_1, \dots, X_n$ are iid with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. 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.

   Solution. First, we note that

   $$S_n^2 = \frac{n}{n-1}(U_n + V_n)$$

   with $U_n = \frac{1}{n}\sum_{i=1}^n X_i^2$ and $V_n = -\bar{X}_n^2$. Since $U_n \to_p E(X^2)$ and $V_n \to_p -\mu^2$ by continuity (see problem 1), we have by Slutsky's lemma that $U_n + V_n \to_p E(X^2) - \mu^2 = \sigma^2$. Since $\frac{n}{n-1} \to_p 1$, further application of Slutsky's lemma leads to the desired conclusion $S_n^2 \to_p \sigma^2$.

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

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

Solution. By the CLT, we have that

$$Z_n = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \to_d N(0,1).$$

We can write

$$T_n = Z_n \frac{\sigma}{\sqrt{S_n^2}}.$$

By part (a) and continuity of $h(t) = \sigma/\sqrt{t}$ at $t \neq 0$, we have that $h(S_n^2) \to_p 1$. Hence, by Slutsky's lemma we have

$$T_n \to_d N(0,1).$$

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

Solution. Let $T_n = \log G_n = \frac{1}{n}\sum_{i=1}^n \log X_i$, which is the mean of iid random variables. By the CLT,

$$\sqrt{n}(T_n - \mu) \to_d N(0, \sigma^2).$$

We have that $G_n = \exp(T_n) = g(T_n)$ with $g(t) = g'(t) = exp(t)$. By the delta method,

$$\sqrt{n}(G_n - e^\mu) \to_d N(0, e^{2\mu}\sigma^2).$$

4. Suppose that $X_1, \dots, X_n$ are iid Uniform$(0,\theta)$ and define $T_n = \max(X_1, \dots, X_n)$. 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?

Solution. From Q13 of problem sheet 1, we know that $Var(T_n)$ is on the order of n^{-2}. To prevent $Var[a_n(T_n - \theta)] \to 0$, we might expect that $a_n = n^1 = n$ is an appropriate scaling factor.

In any case, $P(a_n(T_n - \theta) \le t) = P(T_n \le \theta + t/a_n) = \left(1 + \frac{t/\theta}{a_n}\right)^n$, where the probability is derived as in Q13 of problem sheet 1. It is now evident that for $a_n = n$, as $n \to \infty$,

$$P(a_n(T_n - \theta) \le t) = \left(1 + \frac{t/\theta}{n}\right)^n \to e^{t/\theta}$$

for $t < 0$ and $P(a_n(T_n - \theta) \le t) = 1$ for all $t \ge 0$. That this sequence is supported on the negative reals follows immediately from noting that $P(T_n < \theta) = 1$ since $0 < X_i < \theta$ for each i.

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.

Solution. We will make use of the following identity

$$\begin{aligned} P(|T_n - \theta| < \epsilon) &= P(-\epsilon < T_n - \theta < \epsilon) \\ &= P(T_n - \theta < \epsilon) - P(T_n - \theta \le -\epsilon) \\ &= P(\sqrt{n}(T_n - \theta) < \sqrt{n}\epsilon) - P(\sqrt{n}(T_n - \theta) \le -\sqrt{n}\epsilon). \end{aligned}$$

We want to show that for any $\delta > 0$, there exists n_0 such that for $n > n_0$ we have

$$P(|T_n - \theta| < \epsilon) > 1 - \delta.$$

Let $z > 0$ be such that $\Phi(z/\sigma) - \Phi(-z/\sigma) = 1 - \delta/2$, where $\Phi(t)$ denotes the standard normal cdf. Whenever $\sqrt{n}\epsilon > z \Leftrightarrow n > (z/\epsilon)^2$, we find (by the identity above) that

$$P(|T_n - \theta| < \epsilon) \geq P(\sqrt{n}(T_n - \theta) < z) - P(\sqrt{n}(T_n - \theta) \leq -z).$$

By asymptotic normality, the right-hand side converges to $1 - \delta/2$ for this choice of z. By definition of convergence, there exists a value n_1 such that for any $n > n_1$ the right-hand side is at least $1 - \delta$. Taking $n_0 = \max(n_1, (z/\epsilon)^2)$, we have established that for $n > n_0$ we have

$$P(|T_n - \theta| < \epsilon) > 1 - \delta$$

as desired. This completes the proof.

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)

Solution. We have seen in the previous problem sheet that $\frac{\partial}{\partial\theta} \log f_\theta(x) = \frac{x}{\theta} - \frac{1-x}{1-\theta}$ for $n = 1$. Hence the MLE $\hat{\theta}_n$ solves

$$\sum_{i=1}^n \frac{\partial}{\partial\theta} \log f_\theta(X_i) = \sum_{i=1}^n \frac{X_i}{\theta} - \frac{1 - X_i}{1 - \theta} = \sum_{i=1}^n \frac{X_i - \theta}{\theta(1-\theta)} = 0.$$

Solving for θ we obtain the solution $\hat{\theta}_n = \bar{X}_n$, where $\bar{X}_n$ is the sample mean.
We have also previously shown that $\frac{\partial^2}{\partial\theta^2} \log f_\theta(x) = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2} < 0$ which in turn implies $\hat{\theta}$ is indeed a point of maximum.

(b) Poisson(θ);

Solution. We have

$$f_\theta(x) = \frac{\theta^x e^{-\theta}}{x!}$$
$$\frac{\partial}{\partial\theta} \log f_\theta(x) = \frac{x}{\theta} - 1$$
$$\frac{\partial^2}{\partial\theta^2} \log f_\theta(x) = -\frac{x}{\theta^2}$$

so the MLE solves

$$\sum_{i=1}^n \frac{\partial}{\partial\theta} \log f_\theta(X_i) = \sum_{i=1}^n (\frac{X_i}{\theta} - 1) = 0$$

so that $\hat{\theta}_n = \bar{X}_n$. The second derivative with respect to θ is again negative, so that this is a point of maximum.

(c) Exponential(θ);

Solution. We have

$$\begin{aligned}\log f_\theta(x) &= \log\theta - \theta x\\ \frac{\partial}{\partial\theta}\log f_\theta(x) &= \frac{1}{\theta} - x\\ \frac{\partial^2}{\partial\theta^2}\log f_\theta(x) &= -\frac{1}{\theta^2}\end{aligned}$$

so the MLE solves

$$\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_\theta(X_i) = \sum_{i=1}^{n}\left(\frac{1}{\theta} - X_i\right) = 0$$

and $\hat{\theta}_n = 1/\bar{X}_n$. The second derivative with respect to θ is again negative, so that this is a point of maximum.

7. For the distributions in 6(a-c), find Z such that $\sqrt{n}(\hat{\theta}_n - \theta) \to_d Z$.

Solution. We know from the asymptotic normality of the MLE that in each case

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, I(\theta)^{-1})$$

where $I(\theta)$ is the Fisher information for sample of $n = 1$ individuals. From the previous problem sheet, we know

	Bernoulli	Poisson	Exponential
$I(\theta)$	$1/\theta(1-\theta)$	$1/\theta$	$1/\theta^2$
$I(\theta)^{-1}$	$\theta(1-\theta)$	θ	θ^2

We can also use the CLT directly to verify the distribution for the Bernoulli and Poisson distribution, since the MLE is also the sample mean. For the exponential distribution, the CLT can be applied in tandem with the delta method since the MLE is a differentiable function of the sample mean.

8. For the distributions in 6(a) and 6(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.

Solution. By invariance of the MLE, $\hat{\nu}_n = P_{\hat{\theta}_n}(X_1 = 0)$. For the Bernoulli distribution,

$$\hat{\nu} = 1 - \hat{\theta}$$

and $\sqrt{n}(\hat{\nu} - \nu) \to_d N(0, \theta(1-\theta))$ since $\hat{\nu} - \nu = -(\hat{\theta} - \theta)$.

For the Poisson distribution,

$$\hat{\nu} = e^{-\hat{\theta}}$$

and $\sqrt{n}(\hat{\nu}_n - \nu) = \sqrt{n}(e^{-\hat{\theta}_n} - e^{-\theta}) \to_d N(0, e^{-2\theta}\theta)$ by the delta method with $g(t) = e^{-t}$.

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;

Solution. See, e.g., problem 6(a).

(b) The marginal distribution of the Y_i's;

Solution. Recall from the previous problem sheet that

$$\begin{aligned}
f_\theta(y) &= [\theta p_0]^y (1 - \theta p_0)^{1-y} \\
\frac{\partial}{\partial \theta} \log f_\theta(y) &= \frac{y}{\theta} - p_0 \frac{1-y}{1-\theta p_0} \\
\frac{\partial^2}{\partial \theta^2} \log f_\theta(y) &= -\frac{y}{\theta^2} - p_0^2 \frac{1-y}{(1-\theta p_0)^2} \\
I_n(\theta) &= -nE\left\{\frac{\partial^2}{\partial \theta^2} \log f_\theta(Y)\right\} = \frac{np_0}{\theta(1-\theta p_0)} \\
CRLB_Y &= \frac{\theta(1-\theta p_0)}{np_0}
\end{aligned}$$

Hence, the MLE is the solution to

$$\sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(Y_i) = \sum_{i=1}^n \frac{Y_i}{\theta} - p_0 \frac{1-Y_i}{1-\theta p_0} = 0$$

so that $\hat{\theta}_b = \frac{1}{np_0} \sum_{i=1}^n Y_i$. This is indeed a point of maximum since the second derivative is negative.

(c) The joint distribution of the (Y_i, R_i)'s.

Solution. Recall from the previous problem sheet that

$$\begin{aligned}
f_\theta(y, r) &= p_0^r (1-p_0)^{1-r} \{\theta^y (1-\theta)^{1-y}\}^r \\
\frac{\partial}{\partial \theta} \log f_\theta(y, r) &= r\left[\frac{y}{\theta} - \frac{1-y}{1-\theta}\right] \\
\frac{\partial^2}{\partial \theta^2} \log f_\theta(y, r) &= -r\left[\frac{y}{\theta^2} + \frac{1-y}{(1-\theta)^2}\right] \\
I_n(\theta) &= -nE\left\{\frac{\partial^2}{\partial \theta^2} \log f_\theta(Y, R)\right\} = \frac{np_0}{\theta(1-\theta)} \\
CRLB_{Y,R} &= \frac{\theta(1-\theta)}{np_0}
\end{aligned}$$

Hence, the MLE is the solution to

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f_\theta(Y_i, R_i) = \sum_{i=1}^{n} R_i \left[\frac{Y_i}{\theta} - \frac{1-Y_i}{1-\theta}\right] = 0$$

so that $\hat{\theta}_c = \sum_{i=1}^n Y_i / \sum_{i=1}^n R_i$ or equivalently $\hat{\theta}_c = \sum_{i=1}^n Y_i R_i / \sum_{i=1}^n R_i$ since $Y_i R_i = R_i X_i R_i = X_i R_i = Y_i$. This is indeed a point of maximum since the second derivative is negative.

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.

Solution. Noting that the asymptotic variance of the MLE is the CRLB for $n = 1$ we use calculations of the previous problem below.

In particular, we have that

$$\sqrt{n}(\hat{\theta}_b - \theta) \to_d N(0, \theta(1-\theta p_0)/p_0)$$

and

$$\sqrt{n}(\hat{\theta}_c - \theta) \to_d N(0, \theta(1-\theta)/p_0).$$

The asymptotic relative efficiency of $\hat{\theta}_b$ to $\hat{\theta}_c$ is

$$\frac{\theta(1-\theta p_0)/p_0}{\theta(1-\theta)/p_0} = \frac{1-\theta p_0}{1-\theta} \geq 1$$

with equality iff $p_0 = 1$ (so that X_i is observed with probability 1). Hence, we prefer the MLE $\hat{\theta}_c$ based on the joint distribution of (Y, R) on the basis of the asymptotic relative efficiency.

The asymptotic relative efficiency of $\hat{\theta}_c$ to the "complete data" MLE $\hat{\theta}_a$ is

$$\frac{\theta(1-\theta)/p_0}{\theta(1-\theta)} = \frac{1}{p_0} \geq 1$$

with equality iff $p_0 = 1$ (so that $R_i = 1$ and X_i is observed with probability 1).

Roughly speaking, a sample of (Y_i, R_i)s provides only a fraction of the information about θ that direct observation of the X_is would. This fraction is precisely equal to p_0.

R lab: Consistency of the sample median

11. In R, the code below implements the simulation study for $n = 10$ and $\epsilon = 0.1$.

```
set.seed(50011)
result <- logical(length = 1000)
n <- 10
```

```
epsilon <- .1
for(i in 1:1000){
X <- rnorm(n, mean = 0)
m <- median(X)
result[i] <- abs(m - 0) < epsilon
}
mean(result)
```

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 your `R` console (or write a script) and then:

(a) Explore how the value of `mean(result)` changes by increasing the value of `n` in this code to, e.g. $n = 30, 50, 100, 200, 500, 1000$.

(b) Referring to the results of your experimentation, comment on whether the sample median appears to be consistent for μ in this setting.

Solution. The purpose of this problem is to obtain a better understanding of convergence in probability and consistency.

Running the above code (including setting the random seed) for each suggested value of n results in the following table

n	10	30	50	100	200	500	1000
$\hat{P}(\lvert m_n - 0\rvert < 0.1)$	0.238	0.340	0.467	0.613	0.764	0.930	0.985

where $\hat{P}(\lvert m_n - 0\rvert < 0.1))$ is the value of `mean(result)`.

Since $\hat{P}(\lvert m_n - 0\rvert < 0.1)) \to 1$ based on the results of this experiment, it suggests the sample median converges to μ in probability (is consistent). To make this argument more compelling, we could repeat the experiment with smaller values of ϵ and different values of μ in the normal distribution.