STA261: Lecture 7

A Central Limit Theorem for the MLE

Alex Stringer

July 25th, 2018

Disclaimer

The materials in these slides are intended to be a companion to the course textbook, *Mathematical Statistics and Data Analysis, Third Edition*, by John A Rice. Material in the slides may or may not be taken directly from this source. These slides were organized and typeset by Alex Stringer.

A big thanks to Jerry Brunner as well for providing inspiration for assignment questions.

License

Sampling distributions

In the previous lectures, we studied the sampling distributions of some estimators.

For example if $X_i \stackrel{IID}{\sim} N(\mu, \sigma^2)$, then

$$\hat{\mu} \sim N\left(\mu, \sigma^2/n\right)$$
$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$$

Do we need to do this work every time we have a new estimator?

Sampling distributions

The examples on the previous slide were MLEs for their respective parameters/distributions. Can we say anything about the sampling distribution of the MLE, in general, for any parameter/distribution?

Let $X_i \overset{IID}{\sim} F_\theta$, let θ_0 represent the "true" value of parameter $\theta \in \Omega$, and let $\hat{\theta} = \text{argmax}\ell(\theta)$ be the MLE, where $\ell(\theta) = \sum_{i=1}^{n} \log f(x_i)|\theta$ is the log-likelihood.

$\hat{\theta}$ is a random variable, because it depends on X.

$\hat{\theta} \sim \ldots?$

Score Statistic

Recall: the *score statistic* or *score vector* or just *score* is the gradient/first derivative of the log-likelihood:

$$S(\theta) = \frac{\partial \ell(\theta)}{\partial \theta}$$

With IID data, $\ell(\theta) = \sum_{i=1}^{n} \ell_i(\theta)$ is a sum of contributions from each datapoint, and so is $S(\theta)$:

$$S(\theta) = \frac{\partial \sum_{i=1}^{n} \ell_i(\theta)}{\partial \theta} = \sum_{i=1}^{n} s_i(\theta)$$

where $s_i(\theta) = \partial \ell_i(\theta)/\partial(\theta)$

Score Statistic

For example, if $X_i \stackrel{IID}{\sim} N(\mu, 1)$, then $\theta = \mu$ and

$$\ell_i(\mu) = \dots$$
$$s_i(\mu) = \dots$$

Score Statistic

If $X_i \overset{IID}{\sim} N(\mu, \sigma^2)$, then $\theta = (\mu, \sigma^2) \in \mathbb{R}^2$ and

$$\begin{aligned} \ell_i(\mu, \sigma^2) &= \dots \\ s_i(\mu) &= \dots \\ s_i(\sigma^2) &= \dots \end{aligned}$$

Score Statistic

$S(\theta)$ is a random variable, because it depends on X.

It's also a sum of independent random variables $s_i(\theta)$. We learned a theorem about sums of independent random variables.

Can we find $E(S)$ and $Var(S)$?

Score Statistic

Theorem: $E(s_i(\theta_0)) = 0$ for $i = 1 \ldots n$, where as above, θ_0 represents the true value of θ.

*Proof: ...

Score Statistic

Theorem: $Var(s_i(\theta_0)) = -i_0(\theta_0)$, (minus) the Fisher Information in a single datapoint.

Recall from before, $i_0(\theta) = -E\left(\partial^2 \ell_i(\theta)/\partial\theta^2\right)$.

Proof: ...

Regularity Conditions

For the above to be true, we need 3 things:

1. $\theta_0 \in \Omega_0$- the true value needs to be an interior point of the parameter space
2. The suppport of F_θ must not depend on θ
3. $\ell(\theta) \in C^3$- the log-likelihood needs to be three times continuously differentiable

These conditions are satisfied for the Normal, Gamma, Bernoulli/Binomial, and many others.

They are not satisfied for, for example, the Uniform.

Central Limit Theorem

With these results in hand, we have our first (of two) Central Limit Theorems for this lecture: the *Central Limit Theorem for the Score Statistic*: under all the above conditions, and IID sampling,

$$\frac{S(\theta_0)}{\sqrt{I(\theta_0)}} \xrightarrow{d} N(0,1)$$

This is merely an application of the original Central Limit Theorem: we have showed that s_i are IID with mean 0 and variance $I(\theta_0)$.

The MLE

What about the MLE?

A single-term Taylor expansion of the score function at the MLE about $\theta = \theta_0$ leads to

$$
\begin{aligned}
(\hat{\theta} - \theta_0) &\approx \frac{S(\theta_0)}{J(\theta_0)} \\
&= \frac{S(\theta_0)}{I(\theta_0)} \times \frac{\frac{1}{n} I(\theta_0)}{\frac{1}{n} J(\theta_0)}
\end{aligned}
$$

The MLE

The LLN gives $\frac{1}{n}(J(\theta_0) - I(\theta_0)) \xrightarrow{p} 0$, so Slutsky gives

$$\frac{\sqrt{\frac{1}{n}I(\theta_0)}}{\sqrt{\frac{1}{n}J(\theta_0)}} \xrightarrow{p} 1$$

hence

$$\sqrt{I(\theta_0)} \times \frac{S(\theta_0)}{J(\theta_0)} = \frac{S(\theta_0)}{\sqrt{I(\theta_0)}} \times \frac{\frac{1}{n}I(\theta_0)}{\frac{1}{n}J(\theta_0)} \xrightarrow{p} Z \times 1$$

with $Z \sim N(0,1)$. Hence $\sqrt{I(\theta_0)}(\hat{\theta} - \theta_0) \xrightarrow{p} N(0,1)$.

The MLE

Hence, the *Central Limit Theorem for the MLE*:

$$\frac{\hat{\theta} - \theta_0}{1/\sqrt{I(\theta_0)}} \xrightarrow{d} N(0,1)$$

We say that *the MLE is asymptotically normally distributed with mean θ_0 and variance $1/I(\theta_0)$*.

We call θ_0 the *asymptotic mean* of $\hat{\theta}$ and $1/I(\theta_0)$ the *asymptotic variance*.

Example

Example: Let $X_i \stackrel{IID}{\sim} N(\mu, \sigma^2)$. Find the MLE and its asymptotic distribution.

Aside: notice how this theorem is *exact* when the data is normally distributed?

Example

Example: Let $X_i \overset{IID}{\sim} Exp(\theta)$ with the $E(X) = \theta$ parametrization.
Find the MLE and its asymptotic distribution.

Example

Example: Let $X_i \overset{IID}{\sim} Exp(\beta)$ with the $E(X) = \frac{1}{\beta}$ parametrization. Find the MLE and its asymptotic distribution.

Example

Example: Let $X_i \overset{IID}{\sim} Gamma(\alpha, \beta)$. Find the MLE and its asymptotic distribution.