Statistical Inference

Lecture 07b

ANU - RSFAS

Last Updated: Thu Apr 20 13:32:18 2017

Evaluating Estimators

- Thus far we have based consideration, typically, on a fixed sample size n. Now let's consider evaluating an estimator when $n \to \infty$.

Definition A sequence of estimators $T_n = T_n(X_1, \ldots, X_n)$ is a consistent sequence of estimators for θ, if for every $\epsilon > 0$ and every $\theta \in \Theta$:

$$\lim_{n \to \infty} P(|T_n - \theta| < \epsilon) = 1$$

Or

$$\lim_{n \to \infty} P(|T_n - \theta| \geq \epsilon) = 0$$

- This is just convergence in probability.

Evaluating Estimators

- Consider an estimator W_n. Then using Chebychev's Inequality we have:

$$P(|W_n - \theta| \geq \epsilon) \leq \frac{E[(W_n - \theta)^2]}{\epsilon^2}$$

$$E[(W_n - \theta)^2] = V(W_n) + [Bias(W_n)]^2$$

Theorem: If W_n is a sequence of estimators of a parameter θ satisfying:

1. $\lim_{n\to\infty} V(W_n) = 0$
2. $\lim_{n\to\infty} Bias(W_n) = 0$

for every $\theta \in \Theta$, then W_n is a consistent sequence of estimators.

MLEs

Theorem B: Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x|\theta)$ and let $\hat{\theta}$ be the MLE of θ. Under regularity conditions of $f(x|\theta)$ and thus $L(\theta|\boldsymbol{x})$ (under appropriate smoothness conditions), we can state:

$$W = \frac{1}{\sqrt{n}} \ell'(\theta|\boldsymbol{x}) \xrightarrow{D} \text{normal}(0, i(\theta))$$

Proof:

$$\frac{\ell'(\theta|\boldsymbol{x})}{\sqrt{n}} = \frac{\sum_{i=1}^{n} \ell'(\theta|x_i)}{\sqrt{n}} = \frac{\frac{n}{n} \sum_{i=1}^{n} \ell'(\theta|x_i)}{\sqrt{n}} = \sqrt{n}\ \bar{\ell}'$$

- $\bar{\ell}'$ is the sample average of the first derivative of the log likelihood.

MLEs - Asymptotics

- We can use the Central Limit theorem! We need to know the mean and variance of $\bar{\ell'}$

$$
\begin{aligned}
E[\bar{\ell'}] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\ell'(\theta|x_i)\right] = E[\ell'(\theta|x_i)] \\
&= \int_{-\infty}^{\infty}\ell'(\theta|x_i)f(x_i|\theta)dx_i \\
&= \int_{-\infty}^{\infty}\frac{\frac{\partial}{\partial\theta}f(x_i|\theta)}{f(x_i|\theta)}f(x_i|\theta)dx_i \\
&= \int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}f(x_i|\theta)dx_i \\
&= \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}f(x_i|\theta)dx_i \\
&= \frac{\partial}{\partial\theta}1 = 0
\end{aligned}
$$

MLEs

$$V[\bar{\ell'}] = \frac{1}{n}V[\ell'(\theta|x_i)] = \frac{1}{n}E[\{\ell'(\theta|x_i)\}^2] = -\frac{1}{n}E[\ell''(\theta|x_i)] = \frac{1}{n}i(\theta)$$

- So let's subtract off the mean and divide by the standard deviation:

$$\frac{(\bar{\ell'} - 0)}{\sqrt{i(\theta)/n}} = \frac{\sqrt{n}(\bar{\ell'} - 0)}{\sqrt{i(\theta)}} = \frac{\frac{\ell'(\theta|\boldsymbol{x})}{\sqrt{n}}}{\sqrt{i(\theta)}} \xrightarrow{D} \mathrm{normal}(0, 1)$$

- So

$$\frac{\ell'(\theta|\boldsymbol{x})}{\sqrt{n}} \xrightarrow{D} \mathrm{normal}(0, i(\theta))$$

MLEs

Theorem: Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x|\theta)$. Let $\hat{\theta}$ be the MLE of θ. Under regularity conditions of $f(x|\theta)$ and thus $L(\theta|\boldsymbol{x})$ (under appropriate smoothness conditions), we have:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathrm{normal}(0, i(\theta)^{-1})$$

MLEs

Proof:

- Conduct a Taylor's series expansion of the first derivative of the log likelihood around the true value θ_0:

$$\ell'(\theta|\boldsymbol{x}) = \ell'(\theta_0|\boldsymbol{x}) + (\theta - \theta_0)\ell''(\theta_0|\boldsymbol{x}) + \cdots$$

- Substitute the MLE for θ:

$$\ell'(\hat{\theta}|\boldsymbol{x}) = \ell'(\theta_0|\boldsymbol{x}) + (\hat{\theta} - \theta_0)\ell''(\theta_0|\boldsymbol{x}) + \cdots$$

- Under the regularity conditions we will ignore higher order terms. Also we know $\ell'(\hat{\theta}|\boldsymbol{x}) = 0$:

$$0 = \ell'(\theta_0|\boldsymbol{x}) + (\hat{\theta} - \theta_0)\ell''(\theta_0|\boldsymbol{x})$$

- Now, replace $\ell''(\theta_0|\boldsymbol{x})$ with its expectation:

$$
\begin{aligned}
0 &= \ell'(\theta_0|\boldsymbol{x}) + (\hat{\theta} - \theta_0)E[\ell''(\theta_0|\boldsymbol{x})] \\
&= \ell'(\theta_0|\boldsymbol{x}) + (\hat{\theta} - \theta_0)E\left[\sum_{i=1}^{n} \ell''(\theta_0|x_i)\right] \\
&= \ell'(\theta_0|\boldsymbol{x}) + (\hat{\theta} - \theta_0)\sum_{i=1}^{n} E\left[\ell''(\theta_0|x_i)\right] \\
&= \ell'(\theta_0|\boldsymbol{x}) + (\hat{\theta} - \theta_0)[-ni(\theta_0)]
\end{aligned}
$$

$$
\Rightarrow (\hat{\theta} - \theta_0) = \frac{-\ell'(\theta_0|\boldsymbol{x})}{-ni(\theta_0)}
$$

- Note: $\frac{1}{n}\ell''(\theta_0|\boldsymbol{x}) \stackrel{\text{LLN}}{\rightarrow} E[\frac{1}{n}\ell''(\theta_0|\boldsymbol{x})] = -i(\theta)$

MLEs

- Multiply through by $\sqrt{n}$:

$$
\begin{aligned}
\sqrt{n}(\hat{\theta} - \theta_0) &= \sqrt{n}\frac{\ell'(\theta_0|\boldsymbol{x})}{ni(\theta_0)} = \sqrt{n}\frac{\ell'(\theta_0|\boldsymbol{x})}{\mathbf{I}(\theta_0)} \\
&= \frac{\frac{1}{\sqrt{n}}\ell'(\theta_0|\boldsymbol{x})}{\frac{1}{n}\mathbf{I}(\theta_0)} = \frac{\frac{1}{\sqrt{n}}l'(\theta_0|\boldsymbol{x})}{i(\theta_0)}
\end{aligned}
$$

- Now we saw that:

$$
W = \frac{1}{\sqrt{n}}\ell'(\theta|\boldsymbol{x}) \xrightarrow{D} \mathrm{normal}(0, i(\theta))
$$

- Since a linear transformation of a normal is normal, we just need the mean and variance:

$$E\left[\frac{W}{i(\theta_0)}\right] = \frac{E[W]}{i(\theta)} = \frac{0}{i(\theta)} = 0$$

$$V\left[\frac{W}{i(\theta_0)}\right] = \frac{V[W]}{i(\theta)^2} = \frac{i(\theta)}{i(\theta)^2} = \frac{1}{i(\theta)}$$

- So we have:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathrm{normal}(0, i(\theta)^{-1})$$

Or

$$\hat{\theta} \dot{\sim} n\left(\theta, \frac{1}{ni(\theta)}\right) = \mathrm{normal}(\theta, \mathbf{I}(\theta)^{-1})$$

Delta Method

Theorem (See Rice 4.6): Let Y_n be a sequence of random variables such that:

$$\sqrt{n}(Y_n - \theta) \xrightarrow{D} \mathrm{normal}(0, \sigma^2)$$

- For a given function g and a specific value θ, suppose that $g'(\theta)$ exists and is not 0, then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{D} \mathrm{normal}(0, \sigma^2[g'(\theta)]^2)$$

MLEs

- We can extend the theorem to functions $\tau(\theta)$:

Theorem: Let $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} f(x|\theta)$. Let $\hat{\theta}$ be the MLE of θ and let $tau(\theta)$ be a continuous function of θ. Under regularity conditions (i.e. under appropriate smoothness conditions) of $f(x|\theta)$ and thus $L(\theta|\boldsymbol{x})$, we have:

$$\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{D} \text{normal}(0, \nu(\theta))$$

- Where $\nu(\theta) = \frac{[\tau'(\theta)]^2}{i(\theta)}$ is the Cramer-Rao lower bound for a single data point.

Or

$$\tau(\hat{\theta}) \dot{\sim} \text{normal}\left(\tau(\theta), \frac{[\tau'(\theta)]^2}{\mathbf{I}(\theta)}\right)$$

- We can get this result from the Delta method!

MLEs

- So asymptotically, MLEs are:

1. unbiased;
2. achieve the Cramer-Rao lower bound;
3. asymptotically normally distributed.

- Because these estimators achieve (1-3) they are asymptotically efficient!
- We can also note that MLEs are consistent estimators.

Bayesian Asymptotics (Rice 8.6.2) - Rough Idea

- Suppose we have $y_1, \ldots, y_n \sim p(y|\theta)$.
- Let's consider the posterior distribution:

$$\begin{aligned} p(\theta|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\theta)\ p(\theta) \\ &= exp\left[log\ \ p(\boldsymbol{y}|\theta)\right] exp\left[log\ \ p(\theta)\right] \end{aligned}$$

- As $n \to \infty$ the posterior is dominated by the likelihood. When n is large the prior is nearly constant.

$$p(\theta|\boldsymbol{y}) \propto exp\left[log\ \ p(\boldsymbol{y}|\theta)\right]$$

- Thus to an approximation we have the following:

$$
\begin{aligned}
p(\theta|\boldsymbol{y}) &\propto \propto exp\left[log\ p(\boldsymbol{y}|\theta)\right] \\
&\propto exp\left[\ell(\theta)\right] \\
&\propto exp\left[\ell(\hat{\theta}) + (\theta - \hat{\theta})\ell'(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^2\ell''(\hat{\theta})\right] \\
&\propto exp\left[\frac{1}{2}(\theta - \hat{\theta})^2\ell''(\hat{\theta})\right]
\end{aligned}
$$

- Where: $\hat{\theta}$ is the MLE.
- Note: $\ell(\hat{\theta})$ is a constant.
- Note: $\ell'(\hat{\theta}) = 0$

$$
\begin{aligned}
p(\theta|\boldsymbol{y}) &\propto exp\left[\frac{1}{2}(\theta-\hat{\theta})\ell''(\hat{\theta})\right] \\
&\propto exp\left[\frac{1}{2}(\theta-\hat{\theta})^2\left[-I(\hat{\theta})\right]\right] \\
&\propto exp\left[-\frac{1}{2\left[I(\hat{\theta})\right]^{-1}}(\theta-\hat{\theta})^2\right]
\end{aligned}
$$

- We see that this expression is poportional to a normal distribution. So we have:

$$p(\theta|\boldsymbol{y}) \approx \mathrm{normal}\left(\hat{\theta}, \left[I(\hat{\theta})\right]^{-1}\right)$$