Asymptotic properties of maximum likelihood estimators

Ciaran Evans

Big results so far

- **WLLN:** Under certain conditions, $\overline{X}_n \xrightarrow{p} \mu$

- **CLT:** Under certain conditions, $\dfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} Z \sim N(0, 1)$

These are nice properties! But I don't just want to estimate the mean. Can we say something similar about maximum likelihood estimates in general?

Key results for the MLE

Definition: Let $Y_1, Y_2, ..., Y_n$ be a sample from some distribution with parameter $\theta \in \Theta$. Let $\widehat{\theta}_n$ be an estimator constructed from the sample $Y_1, ..., Y_n$. We say that $\widehat{\theta}_n$ is a **consistent** estimator of θ if, for every $\varepsilon > 0$ and every $\theta \in \Theta$,

$$\lim_{n \to \infty} P_\theta(|\widehat{\theta}_n - \theta| \geq \varepsilon) = 0$$

Key results for the MLE

Let $Y_1, Y_2, \ldots$ be iid from a distribution with probability function $f(y|\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \mathbb{R}^d$ is the parameter(s) we are trying to estimate. Let

$$\ell_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(Y_i|\boldsymbol{\theta})$$

$$\widehat{\boldsymbol{\theta}}_n = \text{argmax}_{\boldsymbol{\theta}} \ell_n(\boldsymbol{\theta})$$

Theorem: Under certain regularity conditions (to be discussed later),

(a) $\widehat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}$

(b) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, v(\boldsymbol{\theta}))$

(where we still need to determine what the variance $v(\boldsymbol{\theta})$ should be!)

Key results for the MLE

Theorem (consistency of the MLE): Let $Y_1, Y_2, \ldots$ be iid from a distribution with parameter $\boldsymbol{\theta}$, and let $\widehat{\boldsymbol{\theta}}_n$ be the MLE constructed from $Y_1, \ldots, Y_n$. Under certain conditions (to be discussed later), $\widehat{\boldsymbol{\theta}}_n$ is a **consistent** estimator of $\boldsymbol{\theta}$.

Key results for the MLE

Theorem (asymptotic normality of the MLE): Let $Y_1, Y_2, \ldots$ be iid from a distribution with parameter $\boldsymbol{\theta}$, and let $\widehat{\boldsymbol{\theta}}_n$ be the MLE constructed from $Y_1, \ldots, Y_n$. Under certain conditions (to be discussed later),

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, v(\boldsymbol{\theta}))$$

(where we still need to determine what the variance $v(\boldsymbol{\theta})$ should be!)

Class activity

Work on the class activity, then we will discuss as a group.

https://sta711-s26.github.io/class_activities/ca_12.html

Class activity

$$Y_1, ..., Y_n \overset{iid}{\sim} Bernoulli(p)$$

$$\ell_n(p) = \log(p)\left(\sum_{i=1}^{n} Y_i\right) + \log(1-p)\left(n - \sum_{i=1}^{n} Y_i\right)$$

$$\ell_n'(p) = \frac{\sum_{i=1}^{n} Y_i}{p} - \frac{\left(n - \sum_{i=1}^{n} Y_i\right)}{1-p}$$

$$\ell_n''(p) = -\frac{\sum_{i=1}^{n} Y_i}{p^2} - \frac{\left(n - \sum_{i=1}^{n} Y_i\right)}{(1-p)^2}$$

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

The expected score

Let Y be a random variable with probability function $f(y|\boldsymbol{\theta})$.

Claim: Under regularity conditions,

$$\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \log f(Y|\boldsymbol{\theta})\right] = \mathbf{0}$$

Fisher information

Let Y be a random variable with probability function $f(y|\boldsymbol{\theta})$.

Definition: The **Fisher information** for a single sample Y, denoted $\mathcal{I}_1(\boldsymbol{\theta})$, is defined as

$$\mathcal{I}_1(\boldsymbol{\theta}) = Var\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log f(Y|\boldsymbol{\theta})\right)$$

Claim: Under regularity conditions,

$$Var\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log f(Y|\boldsymbol{\theta})\right) = -\mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log f(Y|\boldsymbol{\theta})\right]$$

Fisher information

Claim: Under regularity conditions,

$$Var\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log f(Y|\boldsymbol{\theta})\right) = -\mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log f(Y|\boldsymbol{\theta})\right]$$

Asymptotic normality: proof approach

We will sketch the proof in the case that $\theta \in \mathbb{R}$. Let $\ell_n'(\theta) = \frac{d}{d\theta}\ell_n(\theta)$, $\ell_n''(\theta) = \frac{d^2}{d\theta^2}\ell_n(\theta)$

Using a Taylor expansion of ℓ_n' around θ:

$$\widehat{\theta}_n - \theta \approx \frac{\ell_n'(\theta)}{-\ell_n''(\theta)}$$

Asymptotic normality: proof approach

$$\sqrt{n}(\widehat{\theta}_n - \theta) \approx \frac{\frac{1}{\sqrt{n}}\ell_n'(\theta)}{-\frac{1}{n}\ell_n''(\theta)}$$