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

range of values for Θ (for example, $p \in [0,1]$ for Bernoulli(p))

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 $\theta \in \Theta$,

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

← convergence in probability

probability if Θ is true parameter

i.e., regardless of true value of Θ, $\hat{\Theta}_n \xrightarrow{p} \Theta$

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

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

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

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

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

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

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

Proofs make use of WLLN and CLT!

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 $\boldsymbol{\theta}$.

Idea:

1. $\frac{1}{n} \ell_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log f(Y_i | \theta) \xrightarrow[\text{(WLLN)}]{p} \mathbb{E}[\log f(Y_i | \theta)]$

2. Let θ_0 be true value of θ

   $\mathbb{E}[\log f(Y_i | \theta)] < \mathbb{E}[\log f(Y_i | \theta_0)]$ if $\theta \neq \theta_0$

   (expected log-likelihood is maximized at true θ_0!)

3. $\frac{1}{n} \ell_n(\theta)$ is close to $\mathbb{E}[\log f(Y_i | \theta)]$

   so maximizing $\frac{1}{n} \ell_n(\theta)$ is close to maximizing $\mathbb{E}[\log f(Y_i | \theta)]$

   $\Rightarrow \hat{\theta}_n$ close to θ_0

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!)

Idea: $\hat{\theta}_n$ maximizes $\ell_n(\theta)$. If ℓ_n is differentiable, $\hat{\theta}_n$ solves $\ell_n'(\theta) = \frac{\partial}{\partial\theta} \ell_n(\theta) = 0$

Consider case when $\theta \in \mathbb{R}$. Taylor expansion:

$$\ell_n'(\hat{\theta}_n) \approx \ell_n'(\theta) + \ell_n''(\theta)(\theta - \hat{\theta}_n)$$

($\hat{\theta}_n \xrightarrow{p} \theta$ so $\hat{\theta}_n$ close to θ for big n)

$$0 \approx \ell_n'(\theta) + \ell_n''(\theta)(\theta - \hat{\theta}_n)$$

$$\Rightarrow \quad \hat{\theta}_n - \theta \approx \frac{\ell_n'(\theta)}{-\ell_n''(\theta)}$$

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

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

$\mathbb{E}[Y_i] = p \qquad Var(Y_i) = p(1-p)$

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

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

$$\mathbb{E}[\ell_n'(p)] = \frac{np}{p(1-p)} - \frac{n}{1-p} = 0$$

$$Var(\ell_n'(p)) = \frac{np(1-p)}{(p(1-p))^2} = \frac{n}{p(1-p)}$$

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

$$-\mathbb{E}[\ell_n''(p)] = Var(\ell_n'(p))$$

$$-\mathbb{E}[\ell_n''(p)] = -1\left(-\frac{np}{p^2} - \frac{(n-np)}{(1-p)^2}\right) = \frac{n}{p} + \frac{n}{(1-p)} = \frac{n}{p(1-p)}$$

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

$$Var(\hat{p}) = \frac{1}{n^2} \cdot \sum_i Var(Y_i)$$
$$= \frac{1}{n^2} \cdot n \cdot p(1-p) = \frac{p(1-p)}{n} = \frac{1}{Var(\ell_n'(p))}$$