



Lecture 3: Large Sample Theory

STA261 – Probability & Statistics II

Ofir Harari

Department of Statistical Sciences

University of Toronto



Outline

Asymptotic Normality of Maximum Likelihood Estimators

- Motivating Example

- The Score statistic and the Fisher Information

- The Main Theorem

- Invariance of Maximum Likelihood Estimators and Transformations

Other Large Sample Properties of MLE

- Consistency in Probability

- The Plug-in Principle

Summary and Concluding Remarks



Example: the exponential distribution

- Suppose now that $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$ (with pdf $f(x|\lambda) = \lambda e^{-\lambda x}$, $x \geq 0$)

- Let us find the MLE of λ –

$$\star \mathcal{L}(\lambda) = f(x_1, \dots, x_n | \lambda) = \prod_{i=1}^n f(x_i | \lambda) = \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\}$$

$$\star \ell(\lambda) = n \log \lambda - \lambda \sum_{i=1}^n x_i$$

$$\star \ell'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0 \implies \hat{\lambda} = \frac{1}{\bar{X}} \text{ (surprising?)}$$

yes; because MME for exponential is X instead

$$\star \ell''(\hat{\lambda}) = -\frac{n}{\hat{\lambda}^2} < 0 \implies \text{max}$$

- Let us now explore the sampling distribution of the MLE

- Assume $\lambda = 1$ and draw a histogram of $\hat{\lambda}_{\text{MLE}} = \frac{1}{\bar{X}}$

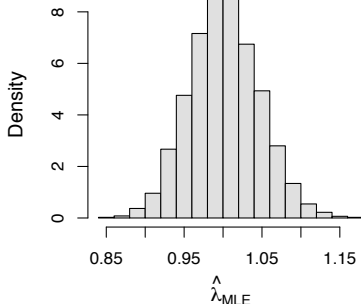


Sampling distribution of $\hat{\lambda}_{MLE}$ by simulation

```
> lambda <- 1
> n <- 20 #sample size
>
> X <- matrix(rexp(n*10000 lambda), ncol=n) # a 10,000 by n matrix of random exponentials
> lambdaHat <- 1/apply(X, 1, mean) # a sample of 10,000 MLEs for lambda
> hist(lambdaHat, freq=FALSE) # plotting the histogram
```

distribution of estimator (for $\sim \text{Exp}$)

Histogram of $1/\bar{X}$ ($n = 500$)





Asymptotic normality

Definition

Let $X_1, \dots, X_n \sim f_\theta$. We say that $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ is *asymptotically normal* with mean θ and variance $\frac{\sigma^2}{n}$ if for all $z \in \mathbb{R}$

$$F_{Z_n}(z) \xrightarrow{n \rightarrow \infty} \Phi(z),$$

where $F_{Z_n}(\cdot)$ is the cdf of $Z_n = \frac{\sqrt{n}}{\sigma} (\hat{\theta}_n - \theta)$ and $\Phi(\cdot)$ is the standard normal cdf.

- Alternatively, write **equivalent to convergence in cdf**

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2) \text{ (converges in distribution)}$$

- Not a new concept: if $\{X_i\}$ are i.i.d r.v.'s with mean μ and variance σ^2 , we know (from the CLT) that

$$Z_n = \sqrt{n} (\bar{X}_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$$

- We expect \bar{X} to be asymptotically normal then – but why would $\frac{1}{\bar{X}} \dots?$
i.e. estimator is asymptotically normal too?



The Score statistic and the Fisher Information

Definition

Let $X_1, \dots, X_n \sim f_\theta$ with $\ell(\theta) = \log f(x_1, \dots, x_n | \theta)$.

1. The *Score* with respect to θ is

$$u(\theta) := \ell'(\theta). \quad \text{statistics: a random variable}$$

2. The *Fisher Information* for θ is

$$\mathcal{I}(\theta) := -\mathbb{E} \{ \ell''(\theta) \}. \quad \text{a scalar}$$

- In the exponential example, we have already calculated

$$u(\lambda) = \ell'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n X_i \quad \text{and} \quad \ell''(\lambda) = -\frac{n}{\lambda^2}$$

- The Fisher information is therefore $\mathcal{I}(\lambda) = -\mathbb{E} \{ \ell''(\lambda) \} = \frac{n}{\lambda^2}$

note n and lambda are both fixed



The Score and the Information (cont.)

- For $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} f_\theta$, denote

$$\mathcal{I}^*(\theta) := -\mathbb{E} \left\{ \frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right\}$$

- the Fisher information of θ based on a single observation

- Note that

$$\begin{aligned} \mathcal{I}(\theta) &= -\mathbb{E} \left\{ \frac{\partial^2 \log f(x_1, \dots, x_n | \theta)}{\partial \theta^2} \right\} = -\mathbb{E} \left\{ \frac{\partial^2 \log \prod_{i=1}^n f(x_i | \theta)}{\partial \theta^2} \right\} \\ &= -\mathbb{E} \left\{ \frac{\partial^2 \sum_{i=1}^n \log f(x_i | \theta)}{\partial \theta^2} \right\} = -\sum_{i=1}^n \mathbb{E} \left\{ \frac{\partial^2 \log f(x_i | \theta)}{\partial \theta^2} \right\} \end{aligned}$$

linearity of derivative

$$= n\mathcal{I}^*(\theta)$$

linearity of expected value



The Score and the Information (cont.)

Proposition

Under some regularity conditions

1. $\mathcal{I}(\theta) = \mathbb{E} [u^2(\theta)]$
2. $\frac{u(\theta)}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}^*(\theta))$ (asymptotically normal)
3. $-\frac{1}{n} \frac{\partial^2 \ell(\theta)}{\partial \theta^2} \xrightarrow{\text{P}} \mathcal{I}^*(\theta)$

Proof:

1. Denoting $\underline{x} = (x_1, \dots, x_n)$, we have

$$\begin{aligned} \mathbb{E}[u(\theta)] &= \int u(\theta) f(\underline{x}|\theta) d\underline{x} = \int \frac{\partial \log f(\underline{x}|\theta)}{\partial \theta} f(\underline{x}|\theta) d\underline{x} = \int \frac{\partial f(\underline{x}|\theta)}{\partial \theta} d\underline{x} \\ &= \frac{\partial}{\partial \theta} \int f(\underline{x}|\theta) d\underline{x} = \frac{\partial}{\partial \theta} (1) = 0. \end{aligned}$$



The Score and the Information (cont.)

We have shown that $\mathbb{E}[u(\theta)] = 0$. Differentiating once again we have

$$0 = \frac{\partial}{\partial \theta} \mathbb{E}[u(\theta)] = \frac{\partial}{\partial \theta} \int \frac{\partial \log f(\underline{x}|\theta)}{\partial \theta} f(\underline{x}|\theta) d\underline{x}$$

product rule $= \int \frac{\partial^2 \log f(\underline{x}|\theta)}{\partial \theta^2} f(\underline{x}|\theta) d\underline{x} + \int \frac{\partial \log f(\underline{x}|\theta)}{\partial \theta} \frac{\partial f(\underline{x}|\theta)}{\partial \theta} d\underline{x}$

chain rule $= \mathbb{E} \left\{ \frac{\partial^2 \log f(\underline{X}|\theta)}{\partial \theta^2} \right\} + \int \left[\frac{\partial \log f(\underline{x}|\theta)}{\partial \theta} \right]^2 f(\underline{x}|\theta) d\underline{x}$

$$= -\mathcal{I}(\theta) + \mathbb{E} \left\{ \left[\frac{\partial \ell(\theta)}{\partial \theta} \right]^2 \right\} = -\mathcal{I}(\theta) + \mathbb{E} [u^2(\theta)] .$$



The Score and the Information (cont.)

2. Since **note the score is a RV i.e. $\sqrt{n} \cdot u / n \sim N(0, \sigma^2)$**

$$\frac{u(\theta)}{\sqrt{n}} = \frac{1}{\sqrt{n}} \frac{\partial \log f(x_1, \dots, x_n | \theta)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \log f(x_i | \theta)}{\partial \theta},$$

it is asymptotically normal (from the CLT). We have already shown that $\mathbb{E}[u] = 0$, and thus

$$\text{Var}[u(\theta)] = \mathbb{E}[u^2(\theta)] = \mathcal{I}(\theta),$$

hence $\text{Var} \left[\frac{u(\theta)}{\sqrt{n}} \right] = \frac{1}{n} \mathcal{I}(\theta) = \mathcal{I}^*(\theta).$

fisher info based on single observation

3. Simple application of the Weak Law of Large Numbers yields

$$-\frac{1}{n} \frac{\partial^2 \ell(\theta)}{\partial \theta^2} = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \log f(x_i | \theta)}{\partial \theta^2} \xrightarrow{P} -\mathbb{E} \left\{ \frac{\partial^2 \log f(x | \theta)}{\partial \theta^2} \right\} = \mathcal{I}^*(\theta).$$



Slutsky's Theorem

- Before we proceed to prove the asymptotic normality of MLEs, we bring (without a proof) the following result:

Slutsky's Theorem

Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of r.v.'s such that $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{\mathcal{P}} c$ (for some constant c), and let $g(\cdot, \cdot)$ be a continuous function. Then

$$g(X_n, Y_n) \xrightarrow{\mathcal{D}} g(X, c).$$

- In particular, if $X_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$ and $Y_n \xrightarrow{\mathcal{P}} c$,

$$X_n Y_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, c^2 \sigma^2).$$

$$X_n Y_n \rightarrow cX$$



Asymptotic normality of MLEs

Theorem

Let X_1, \dots, X_n be a random sample from f_θ and let $\hat{\theta}_n$ denote the maximum likelihood estimator of θ . Under the same regularity conditions as before, $\hat{\theta}_n$ is asymptotically normal with mean θ and variance $\mathcal{I}^{-1}(\theta)$.

“Proof”:

- When deriving the Newton-Raphson iterations, we showed that

$$\hat{\theta}_{\text{MLE}} - \theta \approx -\frac{\ell'(\theta)}{\ell''(\theta)} = -\frac{u(\theta)}{\ell''(\theta)}.$$

- Write

$$\sqrt{n} \left(\hat{\theta}_{\text{MLE}} - \theta \right) = -\sqrt{n} \cdot \frac{u(\theta)}{\ell''(\theta)} = \frac{\frac{1}{\sqrt{n}} u(\theta)}{-\frac{1}{n} \ell''(\theta)}$$



Asymptotic normality of MLEs (cont.)

“Proof” (cont.):

- So far we have

$$\sqrt{n} \left(\hat{\theta}_{\text{MLE}} - \theta \right) = \frac{\frac{1}{\sqrt{n}} u(\theta)}{-\frac{1}{n} \ell''(\theta)}$$

- In the proposition we proved, we showed that

1. The numerator $\xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}^*(\theta))$

2. The denominator $\xrightarrow{\mathcal{P}} \mathcal{I}^*(\theta)$

- Now seems like the right time to apply Slutsky's Theorem:

$$\sqrt{n} \left(\hat{\theta}_{\text{MLE}} - \theta \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, (\mathcal{I}^*(\theta))^{-2} \mathcal{I}^*(\theta) \right) = \mathcal{N} \left(0, \frac{1}{\mathcal{I}^*(\theta)} \right),$$

- Alternatively, $\hat{\theta}_{\text{MLE}} \sim AN \left(\theta, \frac{1}{n \mathcal{I}^*(\theta)} \right) = AN \left(\theta, \mathcal{I}^{-1}(\theta) \right).$



Back to the exponential example

- We just proved:

$$\hat{\theta}_{\text{MLE}} \sim AN(\theta, \mathcal{I}^{-1}(\theta)) \quad (\text{asymptotically normal})$$

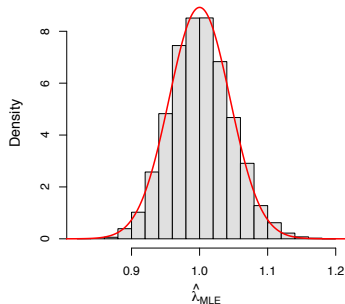
- We have already calculated $\mathcal{I}(\lambda) = \frac{n}{\lambda^2}$
- From the latest Theorem, the distribution of $\hat{\lambda}_{\text{MLE}} = \frac{1}{\bar{X}}$ should be approximately normal, with mean λ and variance $\frac{1}{\mathcal{I}(\lambda)} = \frac{\lambda^2}{n}$



Back to the R simulation

```
> lambda <- 1  
> n <- 500  
> X <- matrix(rexp(n*10000, lambda), ncol=n)  
> lambdaHat <- 1/apply(X, 1, mean)  
> hist(lambdaHat, freq=FALSE)  
> z <- seq(-10, 10, by=.01)  
> lines(z, dnorm(z, mean=lambda, sd=lambda/sqrt(n)))
```

Histogram of $1/\bar{X}$ with normal approximation





Example: Bernoulli distribution

Example

Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Binom}(1, p)$. Find the MLE of p and derive its normal approximation.

Solution:

Let us write the likelihood first –

Note that we are ignoring nCx here since constant does not contribute to likelihood

$$\mathcal{L}(p) = \prod_{i=1}^n f(x_i|p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i},$$

and the log-likelihood –

$$\ell(p) = \sum_{i=1}^n x_i \log p + \left(n - \sum_{i=1}^n x_i \right) \log(1-p).$$



Bernoulli distribution (cont.)

Solution (cont.):

$$\ell(p) = \sum_{i=1}^n x_i \log p + \left(n - \sum_{i=1}^n x_i \right) \log(1 - p).$$

Solving

$$\ell'(p) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1 - p} = 0 \implies (1 - p) \sum_{i=1}^n x_i = \left(n - \sum_{i=1}^n x_i \right) p$$

$$\implies \sum_{i=1}^n x_i = np \implies \hat{p} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

★ Note that

$$\ell''(p) = -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{n - \sum_{i=1}^n x_i}{(1 - p)^2} < 0,$$

therefore it is a maximum.



Bernoulli distribution (cont.)

Solution (cont.):

$$\ell''(p) = -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{n - \sum_{i=1}^n x_i}{(1-p)^2}$$

note fisher information is a function of true population parameter

- The Fisher information is

$$\mathcal{I}(p) = -\mathbb{E} [\ell''(p)] = \frac{\sum_{i=1}^n \mathbb{E}[X_i]}{p^2} + \frac{n - \sum_{i=1}^n \mathbb{E}[X_i]}{(1-p)^2}$$

$$= \frac{np}{p^2} + \frac{n - np}{(1-p)^2} = \frac{n}{p} + \frac{n}{(1-p)} = \frac{n}{p(1-p)},$$

hence

$$\hat{p} \sim AN \left(p, \frac{p(1-p)}{n} \right).$$

- ★ Old news – this is the CLT for Bernoulli r.v.'s

by asymptotic normality of MLE



Invariance of MLEs and transformations

Theorem

Let X_1, \dots, X_n be a sample from f_θ and let $\eta = g(\theta)$ for some function $g(\cdot)$. Then –

1. $\hat{\eta}_{\text{MLE}} = g(\hat{\theta}_{\text{MLE}})$, and
2. If $g(\cdot)$ is differentiable then $\hat{\eta}_{\text{MLE}} \sim AN\left(\eta, [g'(\theta)]^2 \mathcal{I}^{-1}(\theta)\right)$

Proof:

1. Denote $\hat{\eta} = g(\hat{\theta}_{\text{MLE}})$. We need to show that $\hat{\eta} = \hat{\eta}_{\text{MLE}}$, that is:

$f(x_1, \dots, x_n | \eta) \leq f(x_1, \dots, x_n | \hat{\eta})$ For any η . To show that,

$$f(x_1, \dots, x_n | \eta) = \max_{\theta: g(\theta) = \eta} f(x_1, \dots, x_n | \theta) \leq \max_{\theta} f(x_1, \dots, x_n | \theta)$$

how does this work $= f(x_1, \dots, x_n | \hat{\theta}_{\text{MLE}}) = \max_{\theta: g(\theta) = \hat{\eta}} f(x_1, \dots, x_n | \theta)$

$$= f(x_1, \dots, x_n | \hat{\eta}).$$



Invariance of MLEs and transformations (cont.)

Proof (cont.):

2. From the last Theorem, $\hat{\eta}_{\text{MLE}} \sim AN(\eta, \mathcal{I}^{-1}(\eta))$. Now,

$$u(\theta) = \ell'(\theta) = \ell'(\eta)g'(\theta),$$

hence

chain rule: ℓ is a function a function of η

$$\mathcal{I}(\theta) = \mathbb{E}[u^2(\theta)] = [g'(\theta)]^2 \mathbb{E}\left\{[\ell'(\eta)]^2\right\} = [g'(\theta)]^2 \mathcal{I}(\eta),$$

thus

$$\frac{1}{\mathcal{I}(\eta)} = \frac{[g'(\theta)]^2}{\mathcal{I}(\theta)}.$$



Example: MLE for the log-odds

Theorem

Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Binom}(1, p)$. Find the MLE of the log-odds

$$\psi = \log \frac{p}{1-p}$$

and derive its asymptotic sampling distribution.

Solution:

use invariance of MLE to calculate

- We have already calculated $\hat{p}_{\text{MLE}} = \bar{X}$ and $\mathcal{I}(p) = \frac{n}{p(1-p)}$.
- Denote $\theta := g(p) = \log \frac{p}{1-p}$. From the invariance of the MLE,

$$\hat{\theta}_{\text{MLE}} = g(\hat{p}_{\text{MLE}}) = \log \frac{\bar{X}}{1-\bar{X}}.$$



Example: MLE for the log-odds (cont.)

Solution (cont.):

$$g(p) = \log \frac{p}{1-p}$$

- Now,

$$g'(p) = \frac{1-p}{p} \cdot \frac{1 \cdot (1-p) - (-1) \cdot p}{(1-p)^2} = \frac{1}{p(1-p)},$$

and the asymptotic variance of $\hat{\theta}_{\text{MLE}}$ is given by

$$[g'(p)]^2 \mathcal{I}^{-1}(p) = \frac{1}{p^2(1-p)^2} \cdot \frac{p(1-p)}{n} = \frac{1}{np(1-p)}.$$

- In conclusion,

$$\log \frac{\bar{X}}{1-\bar{X}} \sim AN \left(\log \frac{p}{1-p}, \frac{1}{np(1-p)} \right).$$



Consistency of MLEs

intuition: MLE already converges in distribution to normal

- We have shown that $\mathcal{I}(\theta) = n\mathcal{I}^*(\theta)$. From this we gather that –
 1. the Fisher information grows along with the sample size,
 2. the variance of $\hat{\theta}_{\text{MLE}}$ tends to 0 as $n \rightarrow \infty$, and that
 3. $\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta) \sim AN(0, \mathcal{I}^{*-1}(\theta))$ no this is right!!! have to be star

Theorem

If the regularity conditions for the asymptotic normality are satisfied, the MLE is consistent (in probability).

Proof: Recall that the meaning of asymptotic normality is that

$$F_{Z_n}(z) \xrightarrow{n \rightarrow \infty} \Phi(z) \quad \text{for} \quad Z_n = \sqrt{n\mathcal{I}^*(\theta)}(\hat{\theta}_{\text{MLE}} - \theta).$$



Consistency of MLEs (cont.)

- We need to show that for any $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(|\hat{\theta}_{\text{MLE}} - \theta| > \varepsilon \right) = 0,$$

or conversely –

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(|\hat{\theta}_{\text{MLE}} - \theta| \leq \varepsilon \right) = 1.$$

- But

should be $1 - \delta_n$ here

$$\begin{aligned} \mathbb{P} \left(|\hat{\theta}_{\text{MLE}} - \theta| \leq \varepsilon \right) &= \mathbb{P} \left(\left| \sqrt{n\mathcal{I}^*(\theta)}(\hat{\theta}_{\text{MLE}} - \theta) \right| \leq \varepsilon \sqrt{n\mathcal{I}^*(\theta)} \right) \\ &= 2\Phi \left(\varepsilon \sqrt{n\mathcal{I}^*(\theta)} \right) - 1 + \delta_n, \end{aligned}$$

where $\delta_n \rightarrow 0$ as $n \rightarrow \infty$ (why?)

deviation from normality

- Hence

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(|\hat{\theta}_{\text{MLE}} - \theta| \leq \varepsilon \right) = 2 - 1 + 0 = 1.$$



The plug-in principle

- In the Poisson example, we concluded that $\hat{\lambda}_{\text{MLE}} \sim AN\left(\lambda, \frac{\lambda}{n}\right)$
- Likewise, in the Bernoulli example $\hat{p}_{\text{MLE}} \sim AN\left(p, \frac{p(1-p)}{n}\right)$
- In general $\hat{\theta}_{\text{MLE}} \sim AN(\theta, \sigma^2(\theta))$ **is this always true**
- We wish to provide some measure of uncertainty about our estimate of θ
 - A natural candidate: the standard deviation
 - Useless if dependent on θ – the same parameter we wish to estimate

Definition

The standard deviation of an estimator $\hat{\theta}$ of a parameter θ is called the *standard error* of $\hat{\theta}$.



The plug-in principle (cont.)

- The following result is a further application of Slutsky's Theorem:

Theorem

Let $\hat{\theta}_{\text{MLE}}$ be the MLE of θ satisfying the regularity conditions for asymptotic normality, i.e.

$$\hat{\theta}_{\text{MLE}} \sim AN(\theta, \mathcal{I}^{-1}(\theta)).$$

Then for any consistent estimator $\hat{\theta}$ of θ

$$\hat{\theta}_{\text{MLE}} \sim AN(\theta, \mathcal{I}^{-1}(\hat{\theta})).$$

In particular,

$$\hat{\theta}_{\text{MLE}} \sim AN(\theta, \mathcal{I}^{-1}(\hat{\theta}_{\text{MLE}})).$$

pluggin in consistent estimators to
variance does not affect asymptotic
normality of MLE sampling distribution



Example: Bernoulli sample

- Suppose that we flip a coin 100 times and it comes up heads 80 times
- p - the probability of coming up heads
- We have found that $\hat{p}_{\text{MLE}} = \bar{X} = 80/100 = 0.8$
- Moreover, we calculated that $\hat{p}_{\text{MLE}} \sim AN\left(p, \frac{p(1-p)}{n}\right)$
- The latest result states that

$$\hat{p}_{\text{MLE}} \sim AN\left(p, \frac{\bar{X}(1-\bar{X})}{n}\right) = AN\left(p, \frac{0.8 \times 0.2}{100}\right) = AN(p, 0.04^2)$$

plugin estimators to population param

- We say that “the maximum likelihood estimate of p is 0.8, with an estimated standard error of 0.04”.



Summary of Maximum Likelihood Estimation

- Under some regularity conditions

$$\hat{\theta}_{\text{MLE}} \sim AN(\theta, \mathcal{I}^{-1}(\theta)), \quad \text{where} \quad \mathcal{I}(\theta) = -\mathbb{E} [\ell''(\theta)].$$

- If said conditions are met, the MLE is also consistent in probability.
- The MLE of $g(\theta)$ is $g(\hat{\theta}_{\text{MLE}})$. If $g(\cdot)$ is differentiable then

$$g(\hat{\theta}_{\text{MLE}}) \sim AN\left(g(\theta), [g'(\theta)]^2 \mathcal{I}^{-1}(\theta)\right).$$

- When the asymptotic variance of $\hat{\theta}_{\text{MLE}}$ is a function of θ itself, we can substitute $\hat{\theta}_{\text{MLE}}$ for θ . The estimated standard error is then

$$\hat{\sigma}_{\hat{\theta}_{\text{MLE}}} = \mathcal{I}^{-1/2}(\hat{\theta}_{\text{MLE}}).$$

1. MLE sampling distribution
2. MLE is consistent
3. MLE is functional invariant
4. can derive estimated MLE standard error by plug in principle



A note on regularity conditions

- We repeatedly mentioned “regularity conditions” that must be satisfied for MLEs to be asymptotically normal
- A long list of conditions, allowing for the full proof to be carried out
- For example, differentiation and integration need be interchangeable
- $\hat{\theta}_{\text{MLE}}$ must not be a boundary point of the parameter space
- Other conditions concern the differentiability of the likelihood
- When these conditions are violated, $\sqrt{n} \left(\hat{\theta}_{\text{MLE}} - \theta \right)$ will not necessarily converge to a normal distribution
- See handout for an example