STA261 - Module 5

Asymptotic Extensions

Rob Zimmerman

University of Toronto

July 30 - August 1, 2024

Limitations of Finite Sample Sizes

- In almost everything we've done so far, we've assumed a sample $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$ of fixed size n

- We've needed to know the distributions of various statistics of $X_1, X_2, \ldots, X_n$

- This requirement has been very limiting, as the distributions of most statistics don't have closed forms (or are unknown entirely)

- Even the exact distribution of the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$ is only available for a few parametric families

Driving Up the Sample Size

- On the other hand, we have plenty of *limiting* distributions as $n \to \infty$
- Example 5.1:
- Example 5.2:
- Of course, we never have $n = \infty$ in real life
- But if we have the luxury of a very large sample size, the "difference" between the exact distribution and the limiting distribution should (hopefully) be tolerable
- Since the normal distribution is particularly nice, we will milk the CLT for all it's worth

A Review of Standard Limiting Results

- In the following, let $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be sequences of random variables, let X be another random variable, let $x, y \in \mathbb{R}$ be constants, and let $g(\cdot)$ be a continuous function

- Theorem 5.1: If $X_n \xrightarrow{p} X$, then $X_n \xrightarrow{d} X$. If $X_n \xrightarrow{d} x$, then $X_n \xrightarrow{p} x$.

- Theorem 5.2 (**Slutsky's theorem**): If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} y$, then $Y_n \cdot X_n \xrightarrow{d} y \cdot X$ and $X_n + Y_n \xrightarrow{d} X + y$.

- Theorem 5.3 (**Continuous mapping theorem**): If $X_n \xrightarrow{p} X$, then $g(X_n) \xrightarrow{p} g(X)$. If $X_n \xrightarrow{d} X$, then $g(X_n) \xrightarrow{d} g(X)$.

Notation Update

- For the rest of this module, we will accentuate statistics of finite samples with the subscript n (so $\mathbf{X}$ is now $\mathbf{X}_n$, etc.)
- For a generic statistic, we'll write $T_n = T_n(\mathbf{X}_n)$
- If we're talking about a limiting property of a sequence $\{T_n\}_{n \geq 1}$, we'll abuse notation and just write that T_n has that limiting property, when the meaning is clear from context
- Example 5.3:

Two Big Ones

- Theorem 5.4 (**Weak law of large numbers (WLLN)**): Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $\mathbb{E}[X_i] = \mu$. Then

$$\bar{X}_n \xrightarrow{p} \mu.$$

- Theorem 5.5 (**Central limit theorem (CLT)**): Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $\mathbb{E}[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. Then

$$\frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

- The CLT is equivalent to $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$, which is the form we'll be using most often

Poll Time!

On Quercus: Module 5 - Poll 1

Asymptotic Unbiasedness

- As in Module 2, we're interested in estimators of $\tau(\theta)$
- But now we're concerned with their limiting behavious as $n \to \infty$
- For finite n, we insisted that our "best" estimators be unbiased
- In the asymptotic setup, we can relax that slightly
- Definition 5.1: Suppose that $\{W_n\}_{n\geq 1}$ is a sequence of estimators for $\tau(\theta)$. If $\text{Bias}_\theta\left(W_n\right) \xrightarrow{n\to\infty} 0$ for all $\theta \in \Theta$, then $\{W_n\}_{n\geq 1}$ is said to be **asymptotically unbiased** for $\tau(\theta)$.
- Example 5.4:

Consistency

- $\overline{X}_n \xrightarrow{p} \mu$ is the prototypical example of an estimator converging in probability to the "right thing"

- We have a special name for this

- Definition 5.2: A sequence of estimators W_n of $\tau(\theta)$ is said to be **consistent** for $\tau(\theta)$ if $W_n \xrightarrow{p} \tau(\theta)$ for every $\theta \in \Theta$.

- Example 5.5:

Showing Consistency

- Sometimes it's easy to show consistency directly from the definition

- Example 5.6: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$, where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Show that the sample mean $\overline{X}_n$ is consistent for μ.

Showing Consistency

- It's usually easier to use standard limiting results (Slutsky, continuous mapping, etc.) than to go directly from the definition

- Example 5.7: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$, where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Show that the sample variance S_n^2 is consistent for σ^2.

Bringing Back the MSE

- In Module 2, we compared estimators by their MSEs
- To extend that idea to the asymptotic setup, we need a new mode of convergence
- Definition 5.3: Suppose that W_n is a sequence of estimators for $\tau(\theta)$. If $\text{MSE}_\theta\left(W_n\right) \xrightarrow{n\to\infty} 0$ for all $\theta \in \Theta$, then W_n is said to **converge in MSE** to $\tau(\theta)$.
- Example 5.8:

Poll Time!

On Quercus: Module 5 - Poll 2

Convergence in MSE is Already Good Enough

- It turns out that convergence in MSE is strong enough to guarantee consistency

- Theorem 5.6: If W_n is a sequence of estimators for $\tau(\theta)$ that converges in MSE for all $\theta \in \Theta$, then W_n is consistent for $\tau(\theta)$.

Proof.

A Criterion for Consistency

- If we know $\mathbb{E}_\theta [W_n]$ and $\text{Var}_\theta (W_n)$, this next theorem often makes short work out of checking for consistency

- Theorem 5.7: If W_n is a sequence of estimators for $\tau(\theta)$ such that $\text{Bias}_\theta (W_n) \xrightarrow{n\to\infty} 0$ and $\text{Var}_\theta (W_n) \xrightarrow{n\to\infty} 0$ for all $\theta \in \Theta$, then W_n is consistent for $\tau(\theta)$.

Proof.

The Sample Mean is Always Consistent

- Example 5.9: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$, where $\mathbb{E}[X_i] = \mu$. Show that $\overline{X}_n$ is consistent for μ.

The Sample Variance is Always Consistent

- One can (very tediously) show that if $X_1, X_2, \ldots, X_n$ are a random sample from a distribution with a finite fourth moment, then

$$\text{Var}\left(S_n^2\right) = \frac{\mathbb{E}\left[(X_i - \mathbb{E}[X_i])^4\right]}{n} - \frac{\text{Var}(X_i)^2 (n-3)}{n(n-1)}$$

- Example 5.10: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$, where $\mathbb{E}[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$ and $\mathbb{E}\left[X_i^4\right] < \infty$. Show that S_n^2 is consistent for σ^2.

Choosing Among Consistent Estimators

- Consistency is practically the bare minimum we can ask for from a sequence of estimators
- There are usually plenty of sequences that are consistent for $\tau(\theta)$
- Which one should we use?
- It's tempting to go with whichever has the lowest variance for fixed n, but that would rule out a lot of fine estimators
- Example 5.11:

Asymptotic Normality

- There's a much more useful criterion, but first we need an important CLT-inspired definition

- Definition 5.4: Let T_n be a sequence of estimators for $\tau(\theta)$. If there exists some $\sigma^2 > 0$ such that

$$\sqrt{n}[T_n - \tau(\theta)] \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right),$$

  then T_n is said to be **asymptotically normal** with mean $\tau(\theta)$ and **asymptotic variance** σ^2.

- By virtue of the CLT, most unbiased estimators are asymptotically normal

Asymptotic Normality: Examples

- Example 5.12: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bin}(k, p)$. Show that the sample mean $\overline{X}_n$ is asymptotically normal.

Asymptotic Normality: Examples

- Example 5.13: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Exp}\,(\lambda)$. Show that the second sample moment $\overline{X^2}_n$ is asymptotically normal.

Asymptotic Distributions

- More generally, we can talk about the limiting distribution of $\sqrt{n}[T_n - \tau(\theta)]$ even when it's not normal

- Definition 5.5: Suppose that T_n is a sequence of estimators for $\tau(\theta)$. When it exists, the distribution of $\lim_{n\to\infty} \sqrt{n}[T_n - \tau(\theta)]$ is called the **asymptotic distribution** (or **limiting distribution**) of T_n.

- So if T_n is an asymptotically normal sequence of estimators for $\tau(\theta)$ with asymptotic variance σ^2, then its asymptotic distribution is $\mathcal{N}\left(0, \sigma^2\right)$

- Example 5.14:

- We might prefer to speak of the distribution of T_n itself when n is large

Poll Time!

On Quercus: Module 5 - Poll 3

The Delta Method

- If some sequence T_n is asymptotically normal for θ and some function $g(\cdot)$ is nice enough, then the next result gives a remarkably easy method of producing an asymptotically normal sequence of estimators of for $g(\theta)$

- Theorem 5.8 (**Delta method**): Suppose that $\theta \in \Theta \subseteq \mathbb{R}$ and $\sqrt{n}(T_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$. If $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable with $g'(\theta) \neq 0$, then

$$\sqrt{n}[g(T_n) - g(\theta)] \xrightarrow{d} \mathcal{N}\left(0, [g'(\theta)]^2\sigma^2\right).$$

Proof.

The Delta Method: Examples

- Example 5.15: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ where $\mu \in \mathbb{R} \setminus \{0\}$ and $\sigma^2 > 0$. Find the limiting distribution of $1/\overline{X}_n$.

The Delta Method: Examples

- Example 5.16: Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim}$ Bernoulli(θ) where $\theta \in (0,1)$. Find the limiting distribution of $\log\left(1 - \overline{X}_n\right)$.

The Delta Method: Examples

- Example 5.17: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$ where $\mathbb{E}_\theta[X_i] = \theta$ and $\text{Var}_\theta(X_i) = \sigma^2$. If $\tau : \mathbb{R} \to \mathbb{R}$ is continuously differentiable with $\tau'(\theta) \neq 0$, describe the distribution of $\tau(\overline{X}_n)$ as n becomes large.

Back to Choosing Estimators

- We know that when $T_n = \overline{X}_n$, the CLT says that

$$\frac{T_n - \mathbb{E}_\theta\left[T_n\right]}{\sqrt{\mathrm{Var}_\theta\left(T_n\right)}} \xrightarrow{d} \mathcal{N}\left(0, 1\right)$$

- Recall the Fisher information $I_n(\theta) = \mathrm{Var}_\theta\left(S(\theta \mid \mathbf{X}_n)\right)$

- In Module 2, we said that an unbiased estimator W_n of $\tau(\theta)$ was efficient if its variance attained the Cramér-Rao Lower Bound $[\tau'(\theta)]^2/I_n(\theta)$

- We also noticed that if the X_i's were iid, then $I_n(\theta) = nI_1(\theta)$

Asymptotic Efficiency

- So if we could replace the T_n in the CLT statement with a general unbiased and efficient W_n, it would look like

$$\frac{W_n - \tau(\theta)}{\sqrt{[\tau'(\theta)]^2/nI_1(\theta)}} \xrightarrow{d} \mathcal{N}(0,1)$$

- Or equivalently

$$\sqrt{n}[W_n - \tau(\theta)] \xrightarrow{d} \mathcal{N}\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right)$$

- This is not a *result*, but a *condition* that we can demand of our estimators

- Definition 5.6: A sequence of estimators W_n is **asymptotically efficient** for $\tau(\theta)$ if

$$\sqrt{n}[W_n - \tau(\theta)] \xrightarrow{d} \mathcal{N}\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right)$$

Asymptotic Efficiency: Examples

- Example 5.18: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\lambda)$, where $\lambda > 0$. Show that $1/\overline{X}_n$ is asymptotically efficient for λ.

Asymptotic Efficiency: Examples

- Example 5.19: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ Poisson (λ), where $\lambda > 0$. Show that $\overline{X}_n$ is asymptotically efficient for λ.

Large Sample Behaviour for the MLE

- We're ready to see why the MLE is almost always the point estimator of choice when n is large

- To understand this, we need to distinguish between an arbitrary parameter $\theta \in \Theta$ and the true parameter that generated the data, which we will call θ_0

- We'll show that the MLE is asymptotically efficient, under certain "regularity conditions"

Regularity Conditions

- Recall how the Cramér-Rao Lower Bound required some conditions:

- Such conditions are generically referred to as *regularity conditions*, and they're used to rule out various pathological counterexamples and edge cases

- The exact regularity conditions for our next result are quite technical and not worth getting involved with in this course

- Instead, we will go with four *sufficient* regularity conditions that are relatively easy to check, and which are satisfied by many common parametric models

Poll Time!

On Quercus: Module 5 - Poll 4

The MLE is Often Asymptotically Normal

- Theorem 5.9: Let $X_1, X_2, \ldots \overset{iid}{\sim} f_{\theta_0}$, and let $\hat{\theta}_n(\mathbf{X}_n)$ be the MLE of θ_0 based on a sample of size n. Suppose the following regularity conditions hold:

  - Θ is an open interval (not necessarily finite) in $\mathbb{R}$
  - The log-likelihood $\ell(\theta \mid \mathbf{x}_n)$ is three times continuously differentiable in θ
  - The support of f_θ does not depend on θ
  - $I_1(\theta) < \infty$ for all $\theta \in \Theta$

  Then

$$\sqrt{n}[\hat{\theta}_n(\mathbf{X}_n) - \theta_0] \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I_1(\theta_0)}\right).$$

  That is, $\hat{\theta}_n(\mathbf{X}_n)$ is a consistent and asymptotically efficient estimator of θ_0.

Proof (sketch).

A Useful Corollary

- Theorem 5.10: Suppose the hypotheses of Theorem 5.9 hold, and that $\tau : \Theta \to \mathbb{R}$ is continuously differentiable with $\tau'(\theta_0) \neq 0$. Then

$$\sqrt{n}[\tau(\hat{\theta}_n(\mathbf{X}_n)) - \tau(\theta_0)] \xrightarrow{d} \mathcal{N}\left(0, \frac{[\tau'(\theta_0)]^2}{I_1(\theta_0)}\right).$$

  That is, $\tau(\hat{\theta}_n(\mathbf{X}_n))$ is a consistent and asymptotically efficient estimator of $\tau(\theta_0)$.

Asymptotically Efficient MLEs: Examples

- Example 5.20: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$, where $\mu \in \mathbb{R}$ and σ^2 is known. Find the asymptotic distribution of the MLE of μ.

Asymptotically Efficient MLEs: Examples

- Example 5.21: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ Bernoulli (p), where $p \in (0, 1)$. Find the asymptotic distribution of the MLE of p, and then that of $1/p$.

The MLE Isn't Always Asymptotically Normal

- Example 5.22: Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim}$ Unif $(0, \theta)$, where $\theta > 0$. Show that the MLE of θ is not asymptotically normal.

Approximate Tests and Intervals

- We've seen that a lot of statistics are asymptotically normal
- What about test statistics?
- If we're willing to approximate a test statistic (whose exact distribution we might not know for fixed n) by one with a normal distribution, we can perform tests and create intervals that we couldn't have before
- As in Modules 3 and 4, we'll start off with tests and then use the test statistics from those to construct confidence intervals

Wilks' Theorem

- Recall the LRT statistic for testing $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$ was given by $\lambda(\mathbf{X}_n) = \frac{L(\theta_0|\mathbf{X}_n)}{L(\hat{\theta}|\mathbf{X}_n)}$, where $\hat{\theta} = \hat{\theta}(\mathbf{X}_n)$ is the unrestricted MLE of θ based on $\mathbf{X}_n$

- Amazingly, the LRT statistic always converges in distribution to a known distribution, regardless of the statistical model (assuming it's nice enough)

- Theorem 5.11 (**Wilks' theorem**): Let $X_1, X_2, \ldots \overset{iid}{\sim} f_\theta$, where the model satisfies the same regularity conditions as in Theorem 5.9. If we test $H_0 : \theta \in \Theta_0$ versus $H_A : \theta \in \Theta_0^c$ using $\lambda(\mathbf{X}_n)$, then

$$-2\log\left(\lambda(\mathbf{X}_n)\right) \xrightarrow{d} \chi^2_{(1)}$$

under H_0.

Poll Time!

On Quercus: Module 5 - Poll 5

Approximate LRTs: Examples

- Example 5.23: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}\,(p)$, where $p \in (0, 1)$. Construct an approximate size-α LRT of $H_0 : p = p_0$ versus $H_A : p \neq p_0$.

Approximate LRTs: Examples

- Example 5.24: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$, where $\mu \in \mathbb{R}$. Construct an approximate size-α LRT of $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$.

Wald Tests

- Definition 5.7: Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} f_\theta$. For testing $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$, a **Wald test** is a test based on the **Wald statistic**

$$W_n(\mathbf{X}_n) = (\hat{\theta}_n - \theta_0)^2 I_n(\hat{\theta}),$$

where $\hat{\theta}_n = \hat{\theta}_{\text{MLE}}(\mathbf{X}_n)$ is the usual MLE.

- Theorem 5.12: Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} f_\theta$, where the model satisfies the same regularity conditions as in Theorem 5.9. If we test $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$ using $W_n(\mathbf{X}_n)$, then

$$W_n(\mathbf{X}_n) \xrightarrow{d} \chi^2_{(1)}$$

under H_0.

Wald Tests: Examples

- Example 5.25: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(p), where $p \in (0, 1)$. Construct an approximate size-α Wald test of $H_0 : p = p_0$ versus $H_A : p \neq p_0$.

Wald Tests: Examples

- Example 5.26: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$, where $\mu \in \mathbb{R}$. Construct an approximate size-α Wald test of $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$.

Score Tests

- Definition 5.8: Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} f_\theta$. For testing $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$, a **score test** (also called a **Rao test** or a **Lagrange multiplier test**) is a test based on the **score statistic**

$$R_n(\mathbf{X}_n) = \frac{[S_n(\theta_0 \mid \mathbf{X}_n)]^2}{I_n(\theta_0)}.$$

- Theorem 5.13: Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} f_\theta$, where the model satisfies the same regularity conditions as in Theorem 5.9. If we test $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$ using $R_n(\mathbf{X}_n)$, then

$$R_n(\mathbf{X}_n) \xrightarrow{d} \chi^2_{(1)}$$

under H_0.

Score Tests: Examples

- Example 5.27: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$, where $p \in (0, 1)$. Construct an approximate size-α score test of $H_0 : p = p_0$ versus $H_A : p \neq p_0$.

Score Tests: Examples

- Example 5.28: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$, where $\mu \in \mathbb{R}$. Construct an approximate size-α score test of $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$.

The Trinity of Tests

- The LRT, the Wald test, and the score test form the backbone of classical hypothesis testing
- Observe that under H_0, all three tests are asymptotically equivalent (i.e., all three test statistics all converge in distribution to a $\chi^2_{(1)}$)
- For this reason, the three tests are sometimes collectively referred to as the **trinity of tests**
- Although asymptotically equivalent, the speed of convergence to $\chi^2_{(1)}$ can be quite different for each one – for small n, they can be quite different in terms of power and other "small-sample" properties
- One might tell you to reject H_0 while another might not!

Approximate Confidence Intervals

- Using any of the asymptotic tests to test $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$, it's sometimes possible to invert any of the test statistics to obtain an approximate $(1 - \alpha)$-confidence interval for θ

- Out of the three, the LRT is usually the hardest to invert into an actual interval, and the Wald statistic is usually the easiest

- In practice, you can always try to use numerical solvers when the algebra doesn't work

- For Wald and score intervals, the standard recipe is to take the square root of the test statistic and compare it to $\mathcal{N}(0, 1)$

Approximate Confidence Intervals: Examples

- Example 5.29: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$, where $p \in (0, 1)$. Construct an approximate $(1 - \alpha)$-confidence interval for p based on the Wald statistic.

- This confidence interval shows up everywhere in polling (and is a staple of introductory Statistics classes); its half-length is called the **margin of error**

Approximate Confidence Intervals: Examples

- Example 5.30: Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} \text{Bernoulli}\,(p)$, where $p \in (0,1)$. Construct an approximate $(1-\alpha)$-confidence interval for $\log\left(\frac{p}{1-p}\right)$ based on the Wald statistic.

Approximate Confidence Intervals: Examples

- Example 5.31: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$, where $\lambda > 0$. Construct an approximate $(1-\alpha)$-confidence interval for λ based on the Wald statistic.

When the Fisher Information Causes Problems...

- When f_θ is too complicated to allow for exact $(1-\alpha)$-confidence intervals, it's standard practice to use Wald intervals and score intervals
- But there might be another problem:
- In real-life multiparameter models, $I_n(\boldsymbol{\theta})$ is a matrix and is often impossible to work out directly, which makes calculating $I_n(\hat{\boldsymbol{\theta}}_0)$ or $I_n(\hat{\boldsymbol{\theta}})$ futile
- When this happens, people like to swap $I_n(\cdot)$ with $J_n(\cdot)$ in the Wald and score statistics
-
- Moreover, in a famous 1978 paper, Efron and Hinkley showed empirically that $J_n(\hat{\theta})$ is *superior* to $I_n(\hat{\theta})$