Econ 2120: Section 3

Part I - Asymptotics and the Linear Predictor

Ashesh Rambachan

Fall 2018

Outline

Outline

Bias of Least-Squares Estimator

Recall: Under random sampling, least-squares estimator is biased in finite samples for the coefficients of the best linear predictor

Example:

$$Y_i = X_i^3 + \epsilon_i$$

$X_i \sim N(0,1)$ and $\epsilon_i \sim N(0,1)$ with $X_i \perp \epsilon_i$.

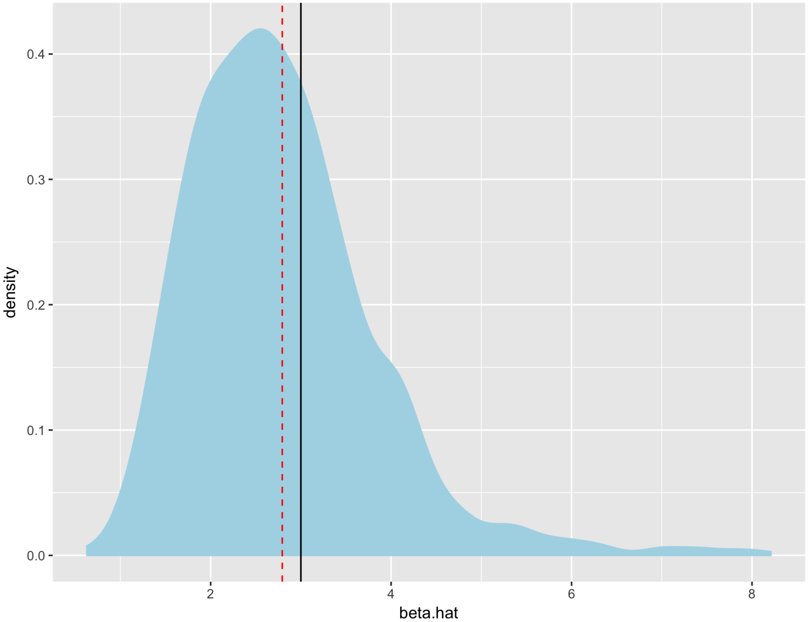
$E^*[Y_i|X_i] = \beta X_i$ with
$\beta = E[Y_i X_i]/E[X_i^2] = E[X_i^4]/E[X_i^2] = 3.$

Simulation:

For $b = 1, \ldots, B$: Draw n pairs of (Y_i^b, X_i^b). Compute $\hat{\beta}^b$ and store it.

$B = 1000$, $n = 30$.

Least-Squares Estimator is biased for β



Bias of Least-Squares Estimator

Recall: Under random-sampling, least-squares estimator is unbiased for

$$\gamma = \arg\min_{\gamma} \sum_{i=1}^{n} [r(x_i) - \gamma x_i]^2$$

i.e. the best linear approximation to the conditional expectation function evaluated at $\{x_i : i = 1, \ldots, n\}$.

Simulation:

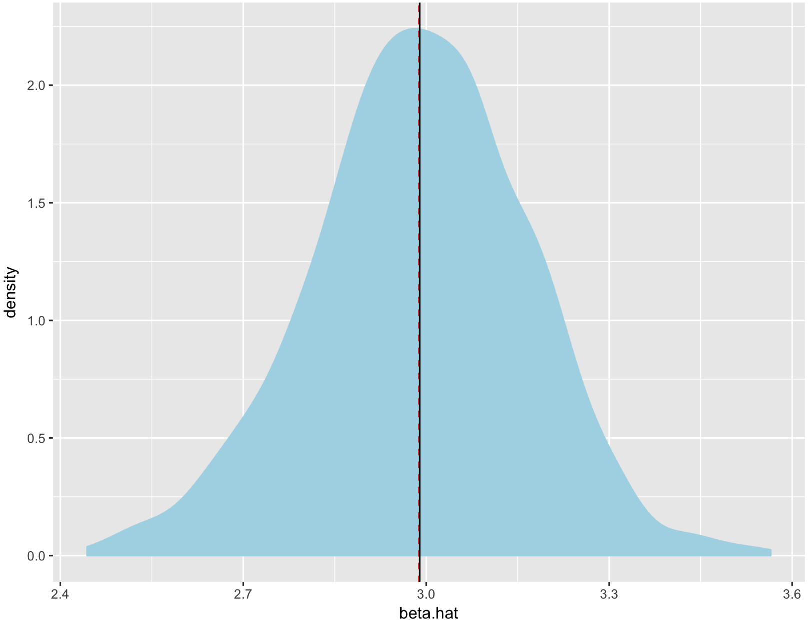
Same model as earlier.

(1) Draw $X_i = x_i$ for $i = 1, \ldots, n$. Construct $r(x_i) = x_i^3$. Compute γ.

(2) For $b = 1, \ldots, B$, draw ϵ_i^b for $i = 1, \ldots, n$. Form $Y_i^b = x_i^3 + \epsilon_i^b$. Construct least-squares estimator $\hat{\beta}^b$ and store it.

$B = 1000$, $n = 30$.

Least-Squares Estimator is unbiased for γ.



Outline

Outline

Convergence in Probability

The sequence of random variables X_n **converges in probability** to a constant α if

$$\lim_{n \to \infty} P(|X_n - \alpha| > \epsilon) = 0$$

for all $\epsilon > 0$.

Convergence in Probability

The sequence of random variables X_n **converges in probability** to a constant α if

$$\lim_{n\to\infty} P(|X_n - \alpha| > \epsilon) = 0$$

for all $\epsilon > 0$.

We write $X_n \xrightarrow{p} \alpha$.

Law of Large Numbers

Theorem

If W_i are i.i.d. with $E[|W_i|] < \infty$, then

$$n^{-1} \sum_{i=1}^{n} W_i \xrightarrow{p} E[W_i].$$

Slutsky Theorem

Also known as **Continuous Mapping Theorem** (CMT).

Theorem

If the sequence of random variables Q_n takes on values in $\mathbb{R}^J$ and $Q_n \xrightarrow{p} \alpha$ and the function $g : \mathbb{R}^J \to \mathbb{R}^M$ is continuous at α, then

$$g(Q_n) \xrightarrow{p} g(\alpha).$$

Outline

Consistent Estimation of β

Observe n realizations of the random vector $(Y_i, X_{i1}, \ldots, X_{iK})$. Let

$$X_i' = (X_{i1}, \ldots, X_{iK}).$$

Assume that (Y_i, X_i') are i.i.d. from some joint distribution.

Consistent Estimation of β

Observe n realizations of the random vector $(Y_i, X_{i1}, \ldots, X_{iK})$. Let

$$X_i' = (X_{i1}, \ldots, X_{iK}).$$

Assume that (Y_i, X_i') are i.i.d. from some joint distribution.

Consider the best linear predictor

$$E^*[Y_i | X_i] = X_i'\beta,$$

and the least-squares estimator

$$\hat{b} = \Big(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i Y_i\Big).$$

Consistent Estimation of β

Observe n realizations of the random vector $(Y_i, X_{i1}, \ldots, X_{iK})$. Let

$$X_i' = (X_{i1}, \ldots, X_{iK}).$$

Assume that (Y_i, X_i') are i.i.d. from some joint distribution.

Consider the best linear predictor

$$E^*[Y_i | X_i] = X_i' \beta,$$

and the least-squares estimator

$$\hat{b} = \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} X_i Y_i \right).$$

The population parameter of interest is $\beta \in \mathbb{R}^K$.

Consistent Estimation

Theorem

$$\hat{b} \xrightarrow{p} \beta$$

as $n \to \infty$.

Proof.

By the LLN,

$$\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \xrightarrow{p} E[X_i X_i'], \quad \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \xrightarrow{p} E[X_i Y_i].$$

Result follows by Slutsky's Theorem and the Continuous Mapping Theorem. □

Some simulations

Same model as earlier:

$Y_i = X_i^3 + \epsilon_i$

$X_i \sim N(0,1)$ and $\epsilon_i \sim N(0,1)$ with $X_i \perp \epsilon_i$.

$E^*[Y_i|X_i] = \beta X_i$ with
$\beta = E[Y_i X_i]/E[X_i^2] = E[X_i^4]/E[X_i^2] = 3$.

Some simulations

Same model as earlier:

$Y_i = X_i^3 + \epsilon_i$

$X_i \sim N(0,1)$ and $\epsilon_i \sim N(0,1)$ with $X_i \perp \epsilon_i$.

$E^*[Y_i|X_i] = \beta X_i$ with

$\beta = E[Y_i X_i]/E[X_i^2] = E[X_i^4]/E[X_i^2] = 3$.

Repeat the simulation as earlier but let n increase from 100 to 5000.

Some simulations

Same model as earlier:

$Y_i = X_i^3 + \epsilon_i$

$X_i \sim N(0,1)$ and $\epsilon_i \sim N(0,1)$ with $X_i \perp \epsilon_i$.

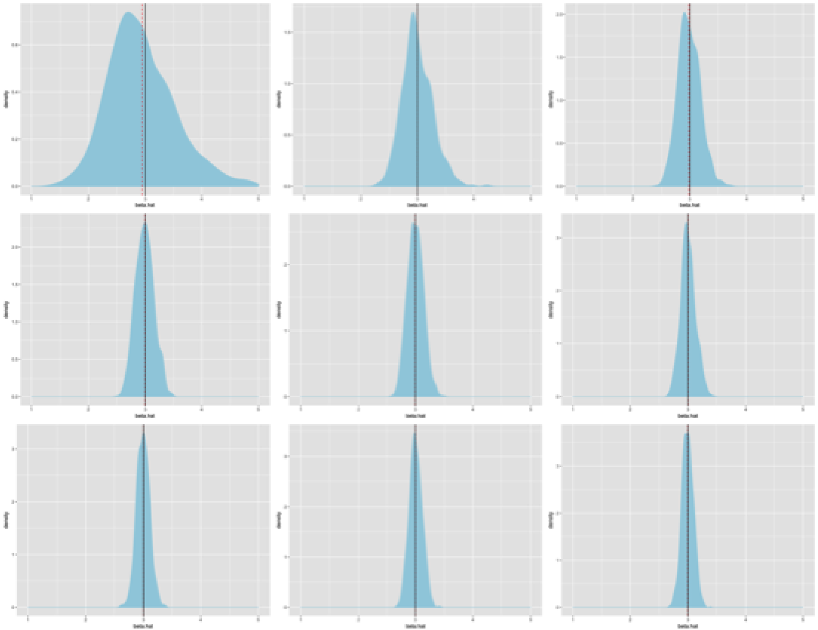
$E^*[Y_i|X_i] = \beta X_i$ with

$\beta = E[Y_i X_i]/E[X_i^2] = E[X_i^4]/E[X_i^2] = 3.$

Repeat the simulation as earlier but let n increase from 100 to 5000.

Look at what happens to the sampling distribution of $\hat{\beta}$. Watch the magic of the LLN unfold before your eyes!

LLN at Work



Outline

Outline

Some simulations

Consistent Estimation of the Linear Predictor Coefficients

- Asymptotics Refresher pt. I
- Consistent Estimation

Asymptotic Distributon of the Least-Squares Estimator

- Asymptotics Review pt. II
- Asymptotic Distribution

Inference

- Confidence Intervals
- Review of Hypothesis Testing
- Hypothesis Testing for the Best Linear Predictor

Homoskedasticity

Convergence in Distribution

Let W be a $K \times 1$ random variable distributed $N(0, \Sigma)$. A sequence of random variables of random variables S_n converges in distribution to $N(0, \Sigma)$ if for an (well-behaved/measurable) subset $A \in \mathbb{R}^K$, we have

$$\lim_{n \to \infty} P(S_n \in A) = P(W \in A).$$

Convergence in Distribution

Let W be a $K \times 1$ random variable distributed $N(0, \Sigma)$. A sequence of random variables of random variables S_n converges in distribution to $N(0, \Sigma)$ if for an (well-behaved/measurable) subset $A \in \mathbb{R}^K$, we have

$$\lim_{n \to \infty} P(S_n \in A) = P(W \in A).$$

We write

$$S_n \xrightarrow{d} W.$$

Central Limit Theorem

Theorem

If the $K \times 1$ random vector G_i are i.i.d. across i with $E[G_i] = 0$ and $Cov(G_i) = \Sigma$, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} G_i = \sqrt{n}\bar{G}_n \xrightarrow{d} N(0, \Sigma),$$

where $\bar{G}_n = n^{-1} \sum_{i=1}^{n} G_i$.

More Slutsky

Theorem

Let S_n be a sequence of $K \times 1$ random variables with $S_n \xrightarrow{N} (0, \Sigma)$. Let Q_n be a sequence of $J \times K$ random variables with $Q_n \xrightarrow{\alpha} \in \mathbb{R}^{J \times K}$. Then

$$Q_n S_n \xrightarrow{d} \alpha \cdot N(0, \Sigma) = N(0, \alpha \Sigma \alpha').$$

Outline

Set-up

As always, define

$$U_i = Y_i - E^*[Y_i|X_i].$$

So,

$$Y_i = X_i'\beta + U_i, \quad E[X_iU_i] = 0.$$

Set-up

As always, define

$$U_i = Y_i - E^*[Y_i|X_i].$$

So,

$$Y_i = X_i'\beta + U_i, \quad E[X_i U_i] = 0.$$

Recall:

$$b = \Big(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i Y_i\Big).$$

Set-up

As always, define

$$U_i = Y_i - E^*[Y_i|X_i].$$

So,

$$Y_i = X_i'\beta + U_i, \quad E[X_iU_i] = 0.$$

Recall:

$$b = \left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_iY_i\right).$$

Sub-in expression for Y_i into expression for b.

Set-up

Algebra magic:

$$b = \beta + \Big(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i U_i\Big).$$

Set-up

Algebra magic:

$$b = \beta + \Big(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i U_i\Big).$$

Re-arrange and multiply by $\sqrt{n}$:

$$\sqrt{n}(b-\beta) = \Big(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\Big)^{-1}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i U_i\Big)$$

Limit Distribution of Least-Squares Estimator

Theorem

$$\sqrt{n}(b - \beta) \xrightarrow{d} N(0, \alpha\Sigma\alpha'),$$

where $\alpha = E[X_i X_i']$ *and* $\Sigma = E[U_i^2 X_i X_i']$.

Limit Distribution of Least-Squares Estimator (proof)

$G_i = X_i U_i$ and so, G_i i.i.d. (random sampling), $E[G_i] = 0$ (orthogonality conditions), $Cov(G_i) = E[U_i^2 X_i X_i']$.

Limit Distribution of Least-Squares Estimator (proof)

$G_i = X_i U_i$ and so, G_i i.i.d. (random sampling), $E[G_i] = 0$ (orthogonality conditions), $Cov(G_i) = E[U_i^2 X_i X_i']$.

CLT:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} G_i \xrightarrow{d} N(0, \Sigma).$$

Limit Distribution of Least-Squares Estimator (proof)

$G_i = X_i U_i$ and so, G_i i.i.d. (random sampling), $E[G_i] = 0$ (orthogonality conditions), $Cov(G_i) = E[U_i^2 X_i X_i']$.

CLT:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} G_i \xrightarrow{d} N(0, \Sigma).$$

LLN + CMT:

$$\left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \xrightarrow{p} E[X_i X_i']^{-1}.$$

Result then follows by Slutsky Theorem.

Limit Distribution of Least-Squares Estimator (proof)

Were you paying attention?

Limit Distribution of Least-Squares Estimator (proof)

Were you paying attention?

CLT requires the existence of $Cov(G_i)$ i.e. existence of second moments.

Limit Distribution of Least-Squares Estimator (proof)

Were you paying attention?

CLT requires the existence of $Cov(G_i)$ i.e. existence of second moments.

We defined $G_i = U_i X_i$. So we need, $E[U_i^2 X_i X_i']$ to be finite.

Limit Distribution of Least-Squares Estimator (proof)

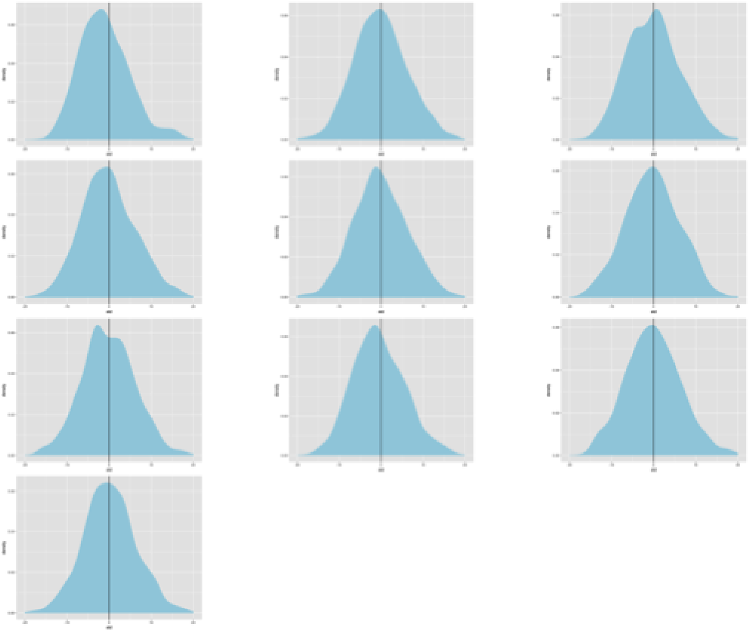
Were you paying attention?

CLT requires the existence of $Cov(G_i)$ i.e. existence of second moments.

We defined $G_i = U_i X_i$. So we need, $E[U_i^2 X_i X_i']$ to be finite.

For this limit distribution argument to work, we need to make an assumption about the 4-th moments of X - $U_i = Y_i - X_i \beta$ and so, $E[U_i^2 X_i X_i']$ is a function of 4th moments.

Some simulations: CLT at work



Outline

Using the Limit Distribution

We have that

$$\sqrt{n}(b - \beta) \xrightarrow{d} N(0, \Lambda),$$

where $\Lambda = \alpha \Sigma \alpha'$.

Using the Limit Distribution

We have that

$$\sqrt{n}(b - \beta) \xrightarrow{d} N(0, \Lambda),$$

where $\Lambda = \alpha \Sigma \alpha'$.

We can use this to create confidence intervals and do hypothesis testing. Just need a consistent estimator of Λ.

Consistent Estimator of Λ

How do we consistently estimate Λ?

Consistent Estimator of Λ

How do we consistently estimate Λ?

Need to consistently estimate

$$\alpha = E[X_i X_i']^{-1}, \quad \text{and} \quad \Sigma = \quad E[U_i^2 X_i X_i'].$$

Consistent Estimator of Λ

How do we consistently estimate Λ?

Need to consistently estimate

$$\alpha = E[X_i X_i']^{-1}, \quad \text{and} \quad \Sigma = \quad E[U_i^2 X_i X_i'].$$

α is easy:

$$\hat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1}.$$

Consistent Estimator of Λ

How do we consistently estimate Λ?

Need to consistently estimate

$$\alpha = E[X_i X_i']^{-1}, \quad \text{and} \quad \Sigma = \quad E[U_i^2 X_i X_i'].$$

α is easy:

$$\hat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1}.$$

For Σ: Plug-in estimates of U_i with $\hat{U}_i = Y_i - X_i' b$. Then,

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \hat{U}_i^2 X_i X_i'.$$

Can show that $\hat{\Sigma} \xrightarrow{p} \Sigma$. See Hayashi Ch. 2 - need additional assumption about 4th moments of X_i.

Consistent Estimator of Λ

Our estimator of Λ is:

$$\hat{\Lambda} = \hat{\alpha}\hat{\Sigma}\hat{\alpha}'$$

Consistent Estimator of Λ

Our estimator of Λ is:

$$\hat{\Lambda} = \hat{\alpha}\hat{\Sigma}\hat{\alpha}'$$

By Slutsky Theorem, $\hat{\Lambda} \xrightarrow{p} \Lambda$.

Outline

Confidence Intervals

Consider linear combinations of coefficients:

$$l'\beta = \sum_{j=1}^{K} l_j \beta_j.$$

Confidence Intervals

Consider linear combinations of coefficients:

$$l'\beta = \sum_{j=1}^{K} l_j \beta_j.$$

Consider the statistic

$$\frac{l'[\sqrt{n}(b-\beta)]}{\sqrt{l'\hat{\Lambda}l}}$$

Confidence Intervals

Consider linear combinations of coefficients:

$$l'\beta = \sum_{j=1}^{K} l_j \beta_j.$$

Consider the statistic

$$\frac{l'[\sqrt{n}(b-\beta)]}{\sqrt{l'\hat{\Lambda} l}}$$

By Slutsky Theorem

$$l'[\sqrt{n}(b-\beta)] \xrightarrow{d} N(0, l'\Lambda l)$$

$$\sqrt{l'\hat{\Lambda} l} \xrightarrow{p} \sqrt{l'\Lambda l}.$$

Confidence Intervals

Theorem

$$\frac{l'[\sqrt{n}(b-\beta)]}{\sqrt{l'\hat{\Lambda}l}} \xrightarrow{d} N(0,1).$$

Confidence Intervals

Theorem

$$\frac{l'[\sqrt{n}(b-\beta)]}{\sqrt{l'\hat{\Lambda}l}} \xrightarrow{d} N(0,1).$$

We refer to the ratio $\frac{l'[\sqrt{n}(b-\beta)]}{\sqrt{l'\hat{\Lambda}l}}$ as an **asymptotic pivot** for $l'\beta$. It depends on the unknown parameters $l'\beta$ but its asymptotic limit distribution is known and doesn't depend on them.

Confidence Intervals

Theorem

$$\frac{l'[\sqrt{n}(b-\beta)]}{\sqrt{l'\hat{\Lambda}l}} \xrightarrow{d} N(0,1).$$

We refer to the ratio $\frac{l'[\sqrt{n}(b-\beta)]}{\sqrt{l'\hat{\Lambda}l}}$ as an **asymptotic pivot** for $l'\beta$. It depends on the unknown parameters $l'\beta$ but its asymptotic limit distribution is known and doesn't depend on them.

We use this asymptotic pivot for **asymptotic** confidence intervals.

Confidence Intervals

The previous theorem immediately gives us that

$$\lim_{n\to\infty} P[l'b - 1.96 \cdot SE \leq l'\beta \leq l'b + 1.96 \cdot SE] = 0.95,$$

where $SE = \sqrt{l'\hat{\Lambda}l}$.

Confidence Intervals

The previous theorem immediately gives us that

$$\lim_{n\to\infty} P[l'b - 1.96 \cdot SE \leq l'\beta \leq l'b + 1.96 \cdot SE] = 0.95,$$

where $SE = \sqrt{l'\hat{\Lambda}l}$.

How do you interpret this?

> As n gets arbitrarily large, the probability that the true parameter lies in the interval (which is random) approaches 95%.

Confidence Intervals

The previous theorem immediately gives us that

$$\lim_{n\to\infty} P[l'b - 1.96 \cdot SE \leq l'\beta \leq l'b + 1.96 \cdot SE] = 0.95,$$

where $SE = \sqrt{l'\hat{\Lambda}l}$.

How do you interpret this?

> As n gets arbitrarily large, the probability that the true parameter lies in the interval (which is random) approaches 95%.

How do you get a 95% confidence interval for a single parameter β_j from this?

Outline

Simple Example

Interested in parameter θ that characterizes the probability distribution of Y. Assume θ is one dimensional for simplicity.

Simple Example

Interested in parameter θ that characterizes the probability distribution of Y. Assume θ is one dimensional for simplicity.

Suppose we have an estimator $\hat{\theta}$ with

$$\hat{\theta} \sim N(\theta, \Omega),$$

which may be coming from an asymptotic approximation. Assume Ω is known (again for simplicity).

Simple Example

Interested in parameter θ that characterizes the probability distribution of Y. Assume θ is one dimensional for simplicity.

Suppose we have an estimator $\hat{\theta}$ with

$$\hat{\theta} \sim N(\theta, \Omega),$$

which may be coming from an asymptotic approximation. Assume Ω is known (again for simplicity).

We have two hypotheses:

$$H_0 : \theta = \theta_0$$

$$H_a : \theta \neq \theta_0.$$

Simple Example

How do we decide between the null and the alternative?

Idea: if $\theta = \theta_0$, then $\hat{\theta}$ should be close to θ_0 with high probability.

We formalize this intuition with a **test statistic**.

Simple Example

How do we decide between the null and the alternative?

Idea: if $\theta = \theta_0$, then $\hat{\theta}$ should be close to θ_0 with high probability.

We formalize this intuition with a **test statistic**.

One candidate:

$$\hat{W} = \frac{(\hat{\theta} - \theta_0)^2}{\Omega}.$$

Test statistics of this form are called **Wald statistics**.

Just the usual t-statistic squared, $t = \frac{\hat{\theta}-\theta}{\sqrt{\Omega}}$.

Simple Example

Under H_0,

$$\hat{W} \sim \chi^2_1.$$

Let's use this to design a rule for rejection H_0.

Rule will take the form: If

$$\hat{W} \leq c \implies \text{fail to reject } H_0$$

$$\hat{W} > c \implies \text{reject } H_0$$

where c is some **critical value**.

Intuition: Large values of $\hat{W}$ are unlikely under H_0.

Simple Example

How do we pick c? We pick c to control **size**.

Size: Probability of rejecting given that H_0 is true.

$$P(\hat{W} > c | H_0) = \alpha$$

Low probability of "Type-I error."

Simple Example

How do we pick c? We pick c to control **size**.

Size: Probability of rejecting given that H_0 is true.

$$P(\hat{W} > c | H_0) = \alpha$$

Low probability of "Type-I error."

We also want tests with high **power**.

Power: Probability of rejecting given that H_a is true.

$$P(\hat{W} > c | H_a)$$

Simple Example

How do we pick c? We pick c to control **size**.

Size: Probability of rejecting given that H_0 is true.

$$P(\hat{W} > c | H_0) = \alpha$$

Low probability of "Type-I error."

We also want tests with high **power**.

Power: Probability of rejecting given that H_a is true.

$$P(\hat{W} > c | H_a)$$

Classical approach: Among tests with the same size α, we want the test with maximal power.

Neyman-Pearson Lemma - classic result in this approach.

Simple Example

Let's write an expression for the power.

Suppose $\theta \neq \theta_0$ and define $\delta = (\theta - \theta_0)/\sqrt{\Omega}$.

We have that the power is

$$P_a[\hat{W} > c] = P_a[\frac{((\hat{\theta} - \theta) - (\theta - \theta_0))^2}{\Omega} > c] \tag{1}$$

$$= P_a[(\frac{\hat{\theta} - \theta_1}{\sqrt{\Omega}} + \frac{\theta_1 - \theta}{\sqrt{\Omega}})^2 > c] \tag{2}$$

$$= P_a[(Z + \delta)^2 > c] \tag{3}$$

where $Z \sim N(0, 1)$. As a function of δ, this is the **power function** of the test.

Plots how the power varies as δ changes. Large values of δ $\implies$ high power - "easy to detect large deviations from the null."

Basis of **power calculations** in experiments.

Simple Example

Suppose that $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$. Then,

$$\Omega \approx n^{-1} V$$

and so,

$$\hat{W} = n \frac{(\hat{\theta} - \theta_0)^2}{V} \overset{approx, H_0}{\sim} \chi_1^2.$$

Set $c = 1 - \alpha$ quantile of χ_k^2. Then, the test has asymptotically correct size.

Simple Example

Under H_a,

$$\hat{W} = \left(\sqrt{n} \frac{\hat{\theta} - \theta}{\sqrt{V}} + \sqrt{n} \frac{\theta - \theta_0}{\sqrt{V}} \right)^2$$

for large n.

Simple Example

Under H_a,

$$\hat{W} = \left(\sqrt{n}\frac{\hat{\theta} - \theta}{\sqrt{V}} + \sqrt{n}\frac{\theta - \theta_0}{\sqrt{V}}\right)^2$$

for large n.

The asymptotic power of the test is 1. Why?

> $\sqrt{n}$ multiplying a term that is not going to zero. This blows up as $n \to \infty$.

Take-away

Good to be aware of the terminology: **size**, **power**.

A lot of areas in econometrics start off with: ”The size of our usual tests under our standard assumptions is too large. What is going on?”

- Weak instruments
  - See 2018 NBER SI methods lecture.
- HAC/HAR inference
  - Size-power tradeoff.

Outline

Hypothesis Testing

We can use this for hypothesis testing as well. Consider the null hypothesis (against the two-sided alternative)

$$H_0 : l'\beta = l'\beta_0.$$

Hypothesis Testing

We can use this for hypothesis testing as well. Consider the null hypothesis (against the two-sided alternative)

$$H_0 : l'\beta = l'\beta_0.$$

Under the null:

$$\frac{l'[\sqrt{n}(b - \beta_0)]}{\sqrt{l'\hat{\Lambda}l}} \xrightarrow{d} N(0,1).$$

Hypothesis Testing

We can use this for hypothesis testing as well. Consider the null hypothesis (against the two-sided alternative)

$$H_0 : l'\beta = l'\beta_0.$$

Under the null:

$$\frac{l'[\sqrt{n}(b-\beta_0)]}{\sqrt{l'\hat{\Lambda}l}} \xrightarrow{d} N(0,1).$$

Choose critical value c s.t. $P(|N(0,1)| > c) = 1 - \alpha$. At this critical value, test statistic has asymptotic size (Type-I error) equal to α and power against the alternative equal to 1.

Wald/F-test

We can also test null hypothesis of the form

$$R'\beta = r,$$

where R is $K \times m$ matrix and r is an $m \times 1$ vector.

Wald/F-test

We can also test null hypothesis of the form

$$R'\beta = r,$$

where R is $K \times m$ matrix and r is an $m \times 1$ vector.

Form the usual Wald statistic:

$$W_n = n \cdot (R'b - r)'[R'\hat{\Lambda}R]^{-1}(R'b - r).$$

Wald/F-test

Claim: If $U_n \xrightarrow{d} N(0, V)$ and $V_n \xrightarrow{p} V$, then $U_n' V_n^{-1} U_n \xrightarrow{d} \chi^2_m$, where $m = dim(U_n)$.

Use the claim to show that

$$W_n \xrightarrow{d} \chi^2_m$$

and so, you can once again form an asymptotic test of the null-hypothesis with the correct asymptotic size.

Outline

Summary so far

Derived the limit distribution of the least-squares estimator using ONLY the assumption of random sampling.

If we make more assumptions, the asymptotic inference becomes more simple.

Summary so far

Derived the limit distribution of the least-squares estimator using ONLY the assumption of random sampling.

If we make more assumptions, the asymptotic inference becomes more simple.

Let's additionally assume:

(1): $E[Y_i|X_i] = X_i'\beta$

(2): $V(Y_i|X_i) = \sigma^2$.

Summary so far

Derived the limit distribution of the least-squares estimator using ONLY the assumption of random sampling.

If we make more assumptions, the asymptotic inference becomes more simple.

Let's additionally assume:

(1): $E[Y_i|X_i] = X_i'\beta$

(2): $V(Y_i|X_i) = \sigma^2$.

Or equivalently,

(1): $E[U_i|X_i] = 0$

(2): $V(U_i|X_i) = E[U_i^2|X_i] = \sigma^2$.

Homoskedasticity

What does this get us? Simplifies the asymptotic variance

$$\Sigma = E[U_i^2 X_i X_i'] = E[E[U_i^2 | X_i] X_i X_i'] = \sigma^2 E[X_i X_i'].$$

So, we have that

$$Avar(b) = \sigma^2 E[X_i X_i']^{-1}.$$

Homoskedasticity

What does this get us? Simplifies the asymptotic variance

$$\Sigma = E[U_i^2 X_i X_i'] = E[E[U_i^2 | X_i] X_i X_i'] = \sigma^2 E[X_i X_i'].$$

So, we have that

$$Avar(b) = \sigma^2 E[X_i X_i']^{-1}.$$

Also simplifies estimation

$$\hat{\sigma}^2 = SSR/n = \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i' b)^2$$

$$\hat{\Lambda} = \hat{\sigma^2} \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1}.$$