Question 2: Confidence Intervals

Setup: $\{X_i\}_{i=1}^N$ i.i.d

$EX_i = M$ unknown

$Var(X_i) = \sigma^2$ known finite

Principle: construct a test statistic, find its distribution, and invert: find the values such that a test doesn't reject

CI for M: example we can use is $\bar{X} = \frac{1}{N}\sum_{i=1}^N X_i$ sample mean.

1: find CI $[a, b]$ such that

$P(M \in [a, b]) \to 1-\alpha$ as $N \to \infty$

Ans: use $t_n = \frac{\bar{X} - M}{\sigma/\sqrt{N}}$ (t statistic / z-statistic)

$E t_n = \frac{1}{\sigma/\sqrt{N}}\left(\frac{1}{N}\sum_i \underbrace{EX_i}_{M} - M\right) = 0$

$Var(t_n) = \frac{1}{(\sigma/\sqrt{N})^2}\frac{1}{N^2}\sum_i Var(X_i) = 1$

Apply CLT: $t_n \xrightarrow{d} Z \sim N(0, 1)$

CDF of $t_n \to$ CDF of Z pointwise

Conclude acceptance region $(-q_{1-\alpha/2}, -q_{\alpha/2})$

where q_α is the α^{th} quantile of Z

$$Pr\left(Z \in \left[-q_{1-\alpha/2}, -q_{\alpha/2}\right]\right) = 1-\alpha$$

so $$Pr\left(t_n \in \left[-q_{1-\alpha/2}, -q_{\alpha/2}\right]\right) \to 1-\alpha \text{ as } N \to \infty$$

So we can rearrange

$$Pr\left(\frac{\frac{1}{N}\sum x_i - M}{\sigma/\sqrt{N}} \in \left[-q_{1-\alpha/2}, -q_{\alpha/2}\right]\right)$$

$$= Pr\left(-\frac{\sigma}{\sqrt{N}} q_{1-\alpha/2} \le \bar{x} - M \le -\frac{\sigma}{\sqrt{N}} q_{1-\alpha/2}\right)$$

$$= Pr\left(\bar{x} + q_{\alpha/2}\frac{\sigma}{\sqrt{N}} \le M \le \bar{x} + \frac{\sigma}{\sqrt{N}} q_{1-\alpha/2}\right)$$

if $\alpha = 0.05$, $q_{\alpha/2} = 1.96$

if $\alpha = 0.01$ $q_{1-\alpha/2} = 2.6$

Known vs unknown σ doesn't change this result, because of Slutsky's theorem

---

2. Suppose goal is $[a,b]$ such that

$$Pr(M \in [a,b]) \ge 1-\alpha$$

This is called a "Probably Approximately Correct" (PAC) statistic. Common in ML and comp sci.

↳ Less rigorous: popular book

Method: Chebyshev's inequality

$$\Pr\left(|Z - \mathbb{E}(Z)| \geq k\sqrt{\mathrm{Var}(Z)}\right) \leq \frac{1}{k^2}$$

Approach: use $t_n = \frac{\bar{x} - M}{\sigma/\sqrt{n}}$: $\mathbb{E}[t_n] = 0$

$\mathrm{Var}(t_n) = 1$

Apply Chebyshev to t_n

$$\Pr\left(|t_n| \geq k\right) \leq \frac{1}{k^2}$$

So from $1-\alpha$ interval, $\alpha = \frac{1}{k^2}$

Set $k = \alpha^{-\frac{1}{2}}$

So test: accept if $|t_n| \leq \alpha^{-\frac{1}{2}}$, reject otherwise.

Invert to get interval

$$[a, b] = \left[\bar{x} - \alpha^{\frac{1}{2}}\sigma/\sqrt{n},\ \bar{x} + \alpha^{\frac{1}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

has $\Pr\left(M \in [a, b]\right) \geq 1 - \alpha$

? Why don't people do this?

Reason is, $[a, b]$ in (2) is much bigger than $[a, b]$ in (1).

Why care? Power: if we want to decide based on $\bar{x}$, want to balance Type I & Type II errors
If interval is large, size ↓, power ↓ also,
so will make fewer type I, more type II errors.
Cause this two ways: Normal distribution ← square root of
has quantiles which grow logarithmically
with α, while Chebyshev bound grows
quadratically; so, especially for small α,
Chebyshev bound is much larger. Ratio of size
goes to ∞ as $\alpha \to 0$

Ex: $\alpha = 0.01$, for normal, this is 2.6
for Chebyshev, $\alpha^{-\frac{1}{2}} = 10$, so interval is
≈ 4 times as wide.

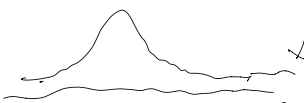
This would be reasonably okay if bound were sharp,
and actual probability was close to $1-\alpha$,
but if inequality is slack

$$\Pr(M \in [a,b]) >> 1-\alpha,$$

then might be better to use refined analysis to tighten.

Part (4) Get a closer bound with more assumptions.
N. finite sample rates if just Finite variance with CLT
CDF converges, but no idea how fast.

And assumption: $E|x_i - M|^3 < \infty$ finite skewness.

Tail pdf  goes to 0 faster than $\frac{1}{x^4}$ (finite skewness)

$$E[x^3] = \int_\alpha^\infty x^3 / x^3 \, dx = A$$

finite skewness

$$E[x^3] = \int_0^A x^3 / x^{4+\varepsilon} \, dx$$

$$= \int_0^\infty \frac{1}{x^{1+\varepsilon}} dx < A$$

So this limitation on probability of large events.

(encounter in financial data, your average is mostly determined by Black Monday, 1/8 of 500 in S&P 500 for decades after, 1 day determines average returns.

Method for CLT: use of convergence in distribution

$$Ef(T_n) \to Ef(S_n)$$ as long as

$$S_n = \frac{1}{N} \sum_{i=1}^{n} \frac{x_i - M}{\sqrt{N}} \qquad T_n = \sum_{i=1}^{n} z_i$$

$$E\frac{x_i - M}{\sqrt{N}} = 0$$

$$\mathrm{Var}\left(\frac{x_i - M}{\sqrt{N}}\right) = \frac{\sigma^2}{N}$$

$$z_i \overset{iid}{\sim} N\left(0, \frac{\sigma^2}{N}\right)$$

Weaker result: for CE, $f(x) = 1\{x \in [a,b]\}$

if we consider $|Ef(T_n) - Ef(S_n)| \le$ bound

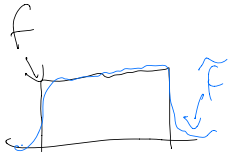
then can use to calculate (a,b) with desired probability

Trick: use Taylor expansion like in proof of CLT.

Let f is not C^3, so we hint,

$\forall \varepsilon > 0, \exists \tilde{f}$ such that $\tilde{f} \in C^3$ and

$$\left| \mathbb{E} f(X) - \mathbb{E} \tilde{f}(X) \right| < \varepsilon$$



$$\left| \mathbb{E} f(S_n) - f(\bar{i}_n) \right|$$

$$\leq \underbrace{\left| \mathbb{E}\left(f(S_n) - \tilde{f}(S_n) \right) \right|}_{< \varepsilon} + \left| \mathbb{E}\left[\tilde{f}(S_n) - \tilde{f}(\bar{i}_n) \right] \right| + \underbrace{\left| \mathbb{E}\left[\tilde{f}(\bar{i}_n) \right] - \mathbb{E}\left[f(\bar{i}_n) \right] \right|}_{< \varepsilon}$$

$$\leq 2\varepsilon + \underbrace{\left| \mathbb{E} \tilde{f}(S_n) - \tilde{f}(\bar{i}_n) \right|}_{\downarrow \text{ as } n \to \infty}$$

$$\leq \frac{16}{\sqrt{n}} \mathbb{E} |x_i - \mu|^3$$

$$\leq \frac{16}{\sqrt{n}} \rho$$

finite

$$\leq \frac{16}{\sqrt{n}} \rho$$

Simply Triangle inequality

$$\Pr(S_n \in [a,b]) \geq \Pr(\bar{T}_n \in [a,b]) - \frac{16\sigma}{\sqrt{N}}\rho$$

make this $\geq 1-\alpha$

But we know distribution of $\bar{T}_n$ exactly, find values $[a,b]$ that do the trick.

$$\frac{\bar{T}_n}{\sigma} \sim N(0,\tfrac{1}{N}) \quad \text{s.t.} \quad a = \sigma\,\Phi^{-1}\left(\tfrac{1}{2}\left(\alpha - \tfrac{16\rho}{\sqrt{N}}\right)\right)$$

$$b = \sigma\,\Phi^{-1}\left(1 - \tfrac{1}{2}\alpha - \tfrac{16\rho}{\sqrt{N}}\right)$$

Then $\Pr(S_n \in [a,b]) \geq 1 - \alpha + \frac{16\rho}{\sqrt{N}} - \frac{16\rho}{\sqrt{N}}$

$$= 1-\alpha$$

So, CI by inverting is

$$\left[\bar{T} - \frac{\sigma}{\sqrt{N}}\Phi^{-1}\left(\tfrac{1}{2}\left(\alpha - \tfrac{16\rho}{\sqrt{N}}\right)\right),\right.$$

$$\left.\bar{T} - \frac{\sigma}{\sqrt{N}}\Phi^{-1}\left(\tfrac{1}{2}\left(\alpha - \tfrac{16\rho}{\sqrt{N}}\right)\right)\right]$$

Use slightly different quantiles of normal distribution where quantile depends on N and gets closer to $\alpha/2$ or $1-\alpha/2$ as $N \to \infty$

- ratio is same as CLT rate of width of CLT interval goes to 1 as $N \to \infty$

This kind of result tells us how accurate CLT limit distributions are.

- Rakhlin, Sridharan, Tewari: "Non-asymptotic analysis for OLS" (PAC)

recently, has been done: Kuchibhotla et al applied this kind of reasoning to OLS in finite samples without assuming normality or exact linearity.

OLS converges to $(\mathbb{E} x x^T)^{-1} \mathbb{E} x y$

---

2. OLS with multiple regressors

$$Y = X_1 \beta + X_2 \beta + \varepsilon$$

$X_1 \in \mathbb{R}^{N \times K_1}$

$X_2 \in \mathbb{R}^{N \times K_2}$

2.1. What if we don't see X_2?

What can we say about β_1?

Sufficient conditions for unbiased estimate

Try estimator: OLS with just X_1

model: $Y = X_1\beta_1 + u$ $\quad \Leftarrow X_2\beta + \varepsilon$

Consistent try OLS: $(X_1'X_1)^{-1}X_1'Y = \hat{\beta}_1$

Writing as

$$\hat{\beta}_1 = \underbrace{(X_1'X_1)^{-1}X_1'X_1}_{I}\beta_1 + (X_1'X_1)^{-1}X_1'u$$

$$= \beta + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon$$

s. $E[\hat{\beta}_1 - \beta] = E(X_1'X_1)^{-1}X_1'X_2\beta_2 + E(X_1'X_1)^{-1}X_1'\varepsilon$

$= 0$ is needed for unbiasedness.

Usual condition, $E[\varepsilon | X_1, X_2] = 0$

implies $E(X_1'X_1)^{-1}X_1'\varepsilon$

$= E\left[(X_1'X_1)^{-1}X_1' \underbrace{E[\varepsilon | X_1, X_2]}_{0}\right)$

Have condition

$E(X_1'X_1)^{-1}X_1'X_2\beta = 0$

when is this true?

Case 1: $E[x_2 | x_1] = 0$

then conditioning on x_1 get $E[(x_1'x_1)^{-1}x_1' E[x_2|x_1]\beta_2]$

$= E[(x_1'x_1)^{-1}x_1' 0] = 0$

or case 2, finite sample orthogonality

$x_1'x_2 = 0$

Geometrically, if they are at right angles, it is finite sample or in expectation.

we are unbiased

In other words if x_1 and x_2 are uncorrelated,

Also finite $\beta_2 = 0$

Related to omitted variables bias, it we set β_2,

Worry about x_2 if it is correlated with x_1

and $\beta_2 \neq 0$

2, A new estimator, show it is unbiased,

a) Regress y on x_1, get residuals $\hat{e}_1$

b) Regress each column of x_2 on x_1, get residuals $\tilde{x}_2$

c) regress $\hat{e}_1$ on $\tilde{x}_2$, get $\tilde{\beta}_2$

Result: a) Residuals $\hat{e}_1 = Y - X_1(X_1'X_1)^{-1}X_1'Y$

$= \underbrace{\left(I - X_1(X_1'X_1)^{-1}X_1'\right)}_{M_1} Y$

M_1 is orthogonal projection onto complement of span of X_1

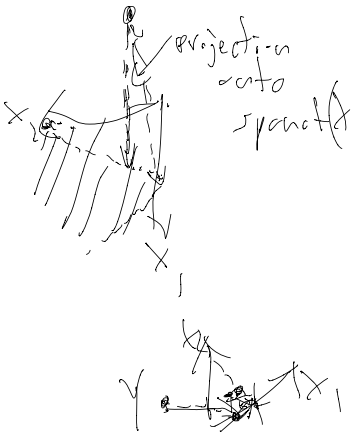
b) Regress X_2: each column X_2^k on X_1

residual $\tilde{X}_2^k = \left(I - X_1(X_1'X_1)^{-1}X_1'\right)X_2^k$

$= M_1 X_2^k$

(in assembling into matrix $\left[\tilde{X}_2^1, \tilde{X}_2^2 \cdots \tilde{X}_2^{k_2}\right]$

$= \tilde{X}_2 = M_1 X_2$



c) Regress $\hat{e}_1$ on $\tilde{X}_2$

Formula $\left(\tilde{X}_2'\tilde{X}_2\right)^{-1}\tilde{X}_2'\hat{e}_1 = \hat{B}_2$

$= \left(X_2'M_1M_1X_2\right)^{-1}X_2'M_1M_1Y$

$$= (X_2' M_1 X_2)^{-1} X_2' M_1 Y$$

by idempotent.

To show bias, use formula $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$

$$\hat{\beta}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 (X_1\beta_1 + X_2\beta_2 + \varepsilon)$$

$M_1 X_1 = 0$ by construction

$$= (X_2' M_1 X_2)^{-1} X_2' M_1 X_2 \beta_2 + (X_2' M_1 X_2)^{-1} X_2' M_1 \varepsilon$$

$$= \beta_2 + (X_2' M_1 X_2)^{-1} X_2' M_1 \varepsilon$$

$$E(\hat{\beta}_2 - \beta_2 \mid X_1, X_2) = E\left((X_2' M_1 X_2)^{-1} X_2' M_1 E(\varepsilon \mid X_1, X_2)\right)$$

$= 0$ under strict exogeneity assumption

so $E\hat{\beta}_2 = \beta_2$ - unbiased

formula $\hat{\beta}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 Y$

is called the partitioned regression formula

It is equal to $\hat{\beta}_2$ component of OLS.

So this estimator is just exactly OLS.

Result known as Frisch Waugh Lowell theorem.

3. Linear Probability Model

$\{Y_i, x_i\}_{i=1}^N \quad Y_i \in \{0, 1\}$

Model: $P(Y_i = 1 | x_i) = x_i'\beta$

(Note: it's continuous, have the careful that this doesn't go above 1 or below 0

Estimating equation: $Y_i = x_i'\beta + \varepsilon_i$

Note $P(Y_i = 1 | x_i)$

$= E[Y_i | x_i) = \underbrace{1 \cdot P(Y_i = 1 | x_i)}_{\text{1}} + \underbrace{0 \cdot P(Y_i = 0 | x_i)}_{0}$

So can apply results on conditional expectations:

If $E[\varepsilon_i | x_i) = x_i'\beta$, then OLS is unbiased.

$\varepsilon_i = Y_i - x_i'\beta$ $\quad E[\varepsilon_i|x_i] = E[Y_i|x_i] - x_i'\beta$

$= x_i'\beta - x_i'\beta$

$= 0$

Why is OLS Not BLUE?

Needed for Gauss Markov homoskedasticity.

$Var(\varepsilon_i|x_i)$

$= E[\varepsilon_i^2|x] - \underbrace{E[\varepsilon|x]^2}_{0}$

$= E((Y_i - x_i'\beta)^2|x] = E[Y^2|x] + E(x'\beta)^2|x)$

$- 2E[Y x'\beta|x]$

N.t. $Y^2 = Y$ since Y is 0 or 1

so this is

$E[Y|x] + E[E[Y|x]^2|x] - 2E[E[Y|x]^2|x]$

$= E[Y|x] - E[Y|x]^2$

$= x'\beta - (x'\beta)^2 = x'\beta(1 - x'\beta)$

is a function of x

$= P_r(Y_i = 1|x)(1 - P_r(Y_i = 1|x))$

(Bernoulli variance formula)

Can use instead GLS.

$$\hat{\beta}^{GLS} = \left(X' \underbrace{Var(\varepsilon|X)^{-1}}_{\Sigma} X\right)^{-1} X' \left(Var(\varepsilon|X)^{-1}\right) y$$

$$\sigma_i^2 = \Sigma_{ii} = (X_i'\beta)(1 - X_i'\beta)$$

$$\Sigma_{ij} = 0 \quad \text{for } i \neq j \text{ by cond.}$$

But this depends on true β.

For this version. Get $\hat{\beta}^{OLS}$ from $Y_i = X_i'\beta + \varepsilon_i$.

Get $\hat{\sigma}_i^2 = X_i'\hat{\beta}(1 - X_i'\hat{\beta}) \to \sigma_i^2$ by consistency

$$\hat{\Sigma} = diag(\hat{\sigma}_i^2)$$

So use $\hat{\beta}^{FGLS} = (X'\hat{\Sigma}^{-1}X)^{-1} X'\hat{\Sigma}^{-1}y$

asymptotically efficient by efficiency result for GLS.

Here assumed correct specification for variance.

Whether LPM is probably wrong

Linear probability Model!