Техническая лекция: Бутстрап

Практическая эконометрика

14 октября 2022 г.

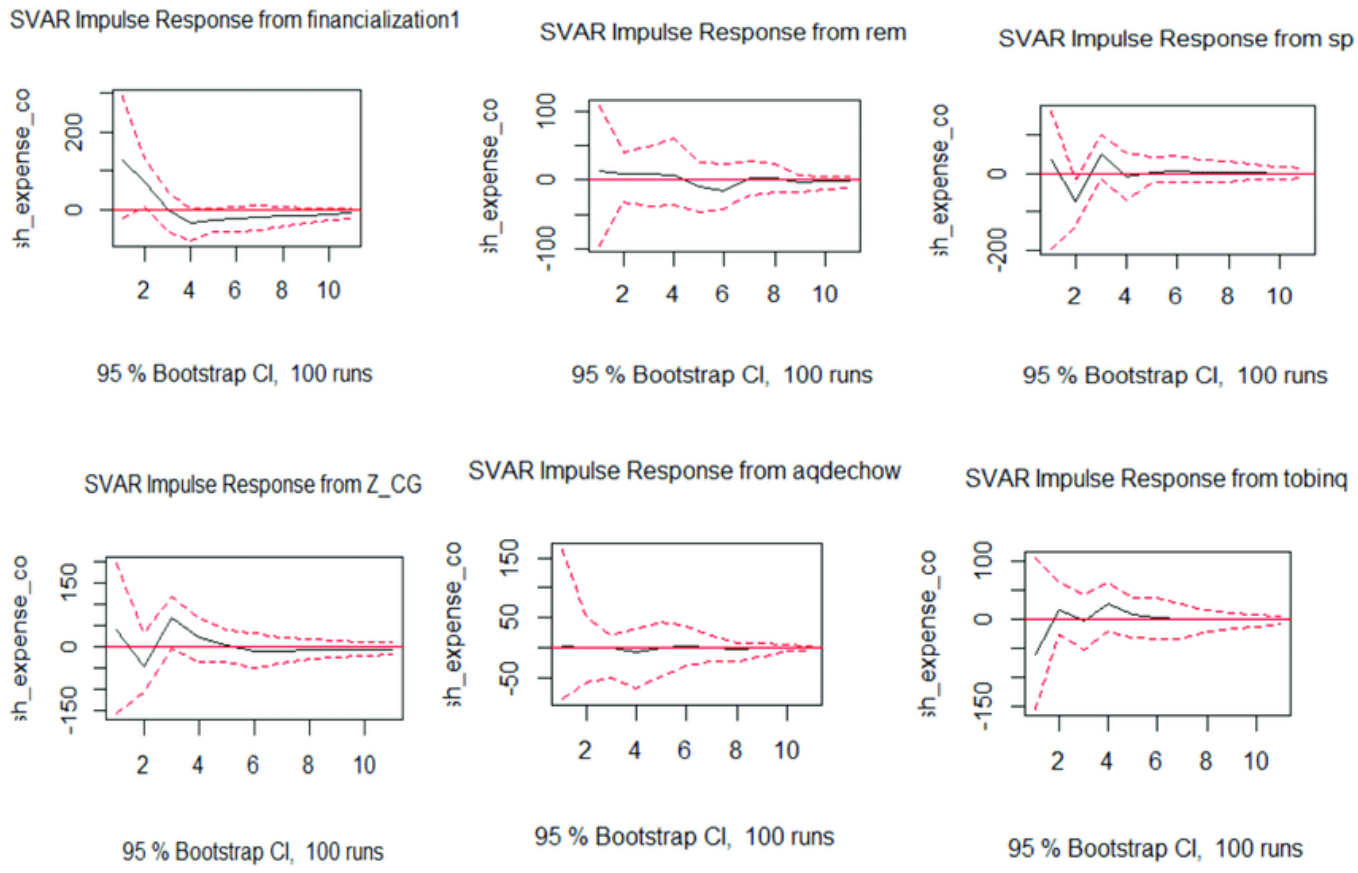
suchkovaolga.91@mail.ru

- Материал этой лекции не войдёт в контрольную 21.10.22
- В слайдах использованы материалы Дмитрия Мухина (Летний семинар-2013), Бориса Демешева и Станислава Анатольева

Пояснения: зачем бутстрап?

- После контрольной содержательно поговорим о гетерогенных тритмент-эффектах
- Один из способов их оценить – «причинный случайный лес» (causal random forest)
- Сначала надо осознать (или вспомнить) деревья решений и просто случайный лес
- Для случайного леса нужен бутстрап
- Поэтому начнём с отступления - бутстрапа.
- Где он уже вам встречался?

В экм-2 вы видели бутстрап-доверительные интервалы для IRF в SVAR



Источник картинки: https://www.researchgate.net/figure/Visualization-of-IRF-for-SVAR-for-Model-4_fig4_362510678

Содержание

- Зачем ещё нужен бутстрап? Пример из жизни
- Решение проблемы 1: бутстрап
- Решение проблемы 2: Дельта-метод
- Что почитать?
- Чем закончилась история

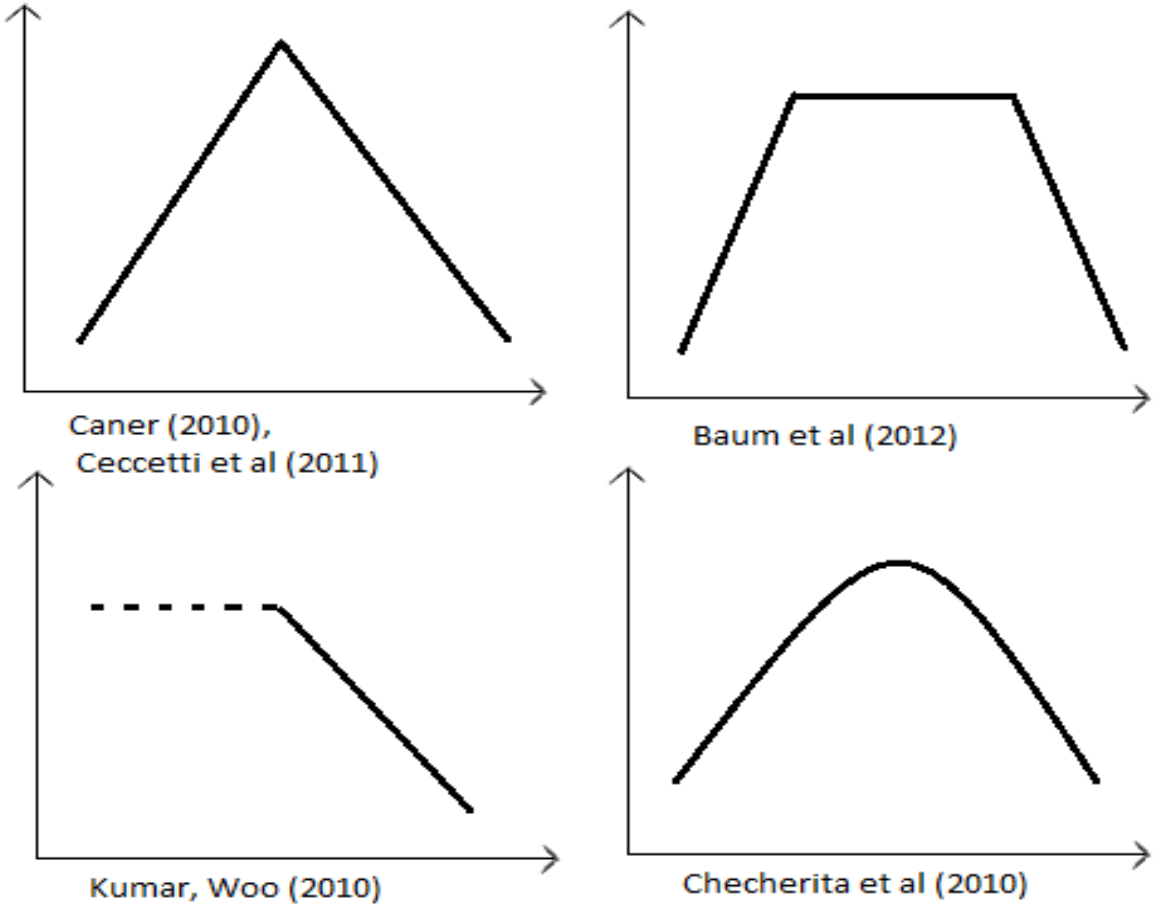
Содержание

- Зачем всё это? Пример из жизни
- Решение проблемы 1: бутстрап
- Решение проблемы 2: Дельта-метод
- Что почитать?
- Чем закончилась история

Пример из жизни (2013 год)

- **Вопрос:** Как государственный долг влияет на динамику валового выпуска?
- **Оценить** для группы европейских стран «точку перелома», начиная с которой уровень долга становится критическим и оказывает негативное влияние на темпы роста выпуска.
- **Проблема:** а как построить доверительный интервал для вершины параболы?

Влияние долговой нагрузки на темпы роста выпуска



- Ось x – долговая нагрузка (долг/ВВП), ось y – темпы роста ВВП

Выбор функциональной формы

1) Пороговая регрессия, предложенная Hansen (1996)

$$g_{it} = \mu_i + \beta_1 debt_{it} I(debt_{it} < c) + \beta_2 debt_{it} I(debt_{it} > c) + u_{it} \quad (1)$$

g_{it} - темп роста выпуска в стране i в период t,
μ_i - фиксированный страновой эффект,
$debt_{it}$ - долг - пороговая переменная,
I - индикаторная функция, гамма – «порог»
u_{it} - случайный шок.

2) Фиктивные переменные (Kumar, Woo (2010))

3) Квадратичная функция (Checherita and Rother (2010))

$$g_{it} = \beta_1 debt_{i,t-1} + \beta_2 debt_{i,t-1}^2 + X_{i,t-1}\beta + u_{it} \quad (2)$$

g_{it} - темп роста выпуска в стране i в период t,
debt - переменная долга,
$X_{i,t-1}$- набор контрольных переменных,
u_{it} - случайный шок.

«Переломная точка» рассчитывается как $-\frac{\beta_1}{2\beta_2}$ (3)

Разброс рассчитанных предельных уровней государственного долга (% от ВВП)

	Развитые страны	Развивающиеся страны	Бедные страны
Pattilo C., Poirson H., Ricci L., 2002		35-40%	
Buiter W. «Fiscal Sustainability»// 2003	60%		
Clements B., Bhattacharya R., Nguyen T.Q., 2003			50% и менее
Caner, Grennes, Koehler-Geib 2010	60%		
Checherita C., Rother Ph., 2010 Cecchetti S., Mohanty M.S., Fabrizio Z., 2011	85-90%		
Reinhart C., Rogoff K., 2010	**85-90%**	60%	
Balázs Égert, 2012	60%	30%	
IMF «Fiscal Monitor: Balancing Fiscal Policy Risks»// 2012	60%	40%	

Результаты и проблема

- Получена оценка «переломной точки» 90-95% для 1990-2007 гг., около 100% для 1990-2010 гг.
- Но почти во всех эмпирических работах разные значения этой точки
- А каков доверительный интервал для неё?
- Проблема: «обычным» способом его построить невозможно.
- Выход?

Содержание

- Зачем всё это? Пример из жизни
- Решение проблемы 1: бутстрап
- Решение проблемы 2: Дельта-метод
- Что почитать?
- Чем закончилась история

«Классическая» статистика

- В статистике «стандартно» есть Теорема вида: *Если выполняются «идеальные условия», то «какая-то формула» сходится к «чему-то известному».*
- Например, КЛММР: если выполняются условия теоремы Г-М и остатки нормальны, то t-расчётное при n стремящемся к бесконечности сходится к N(0,1).
- А что если: Нет теоремы? Не выполнены идеальные условия? n не стремится к бесконечности?

Знакомство с bootstrap

(для начала см. Анатольев, «Квантиль» №3, 2007):

- *Bootstrap (англ.) – петля на заднике ботинка, облегчающая его надевание*.

- Идея метода: имеющаяся выборка – это единственная информация об истинном распределении данных. Поэтому давайте приблизим истинное распределение эмпирическим. То есть «сами себя вытащим».

- Рассмотрим пример

Простой пример (1/3)

- Пусть в выборке всего 2 наблюдения:

$$\begin{pmatrix} x1 \\ y1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} x2 \\ y2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

- Пусть нас интересует θ из модели парной регрессии без константы:

$$y_i = \theta x_i + \varepsilon_i$$

$$\hat{\theta}^{\text{МНК}} = \frac{x1*y1+x2*y2}{(x1)^2+(x2)^2} = \frac{1*0+2*2}{1^2+2^2} = 0{,}8$$

Простой пример (2/3)

- «Вытаскиваем» две пары чисел:

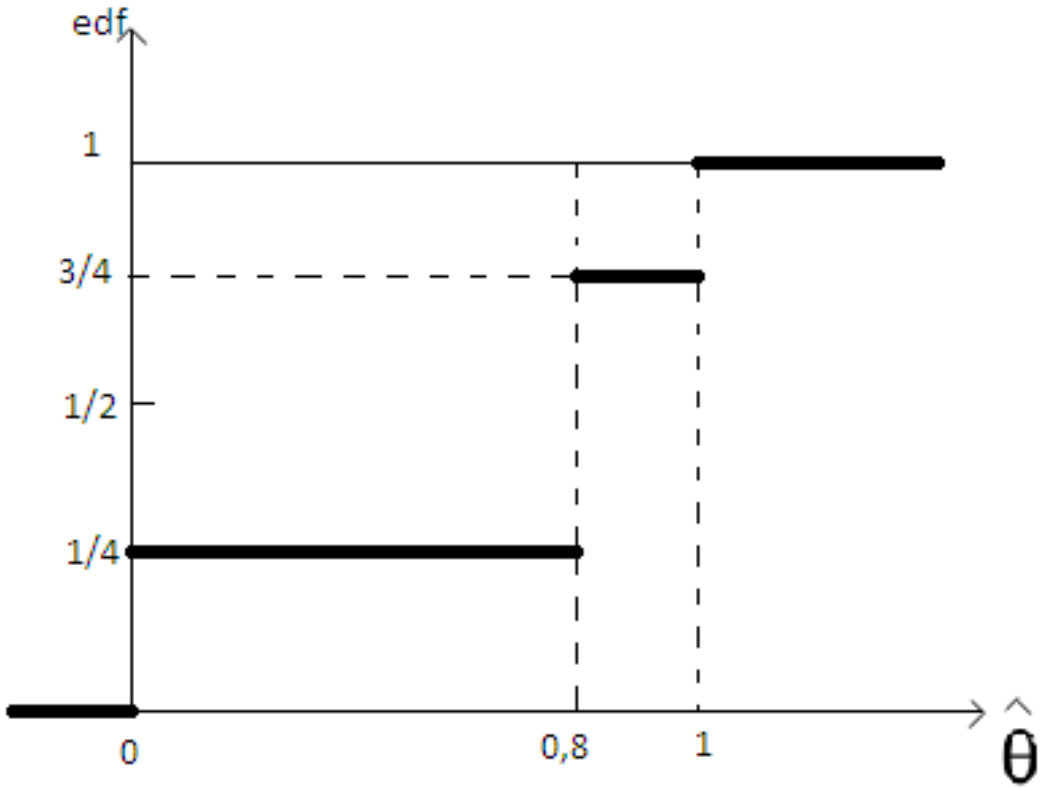
$$\binom{x1^*}{y1^*};\binom{x2^*}{y2^*} = \begin{cases} \binom{1}{0};\binom{1}{0} \text{ с вероятностью } 1/4 \\ \binom{1}{0};\binom{2}{2} \text{ с вероятностью } 1/2 \\ \binom{2}{2};\binom{2}{2} \text{ с вероятностью } 1/4 \end{cases}$$

- Тогда бутстраповская МНК-оценка распределена так:

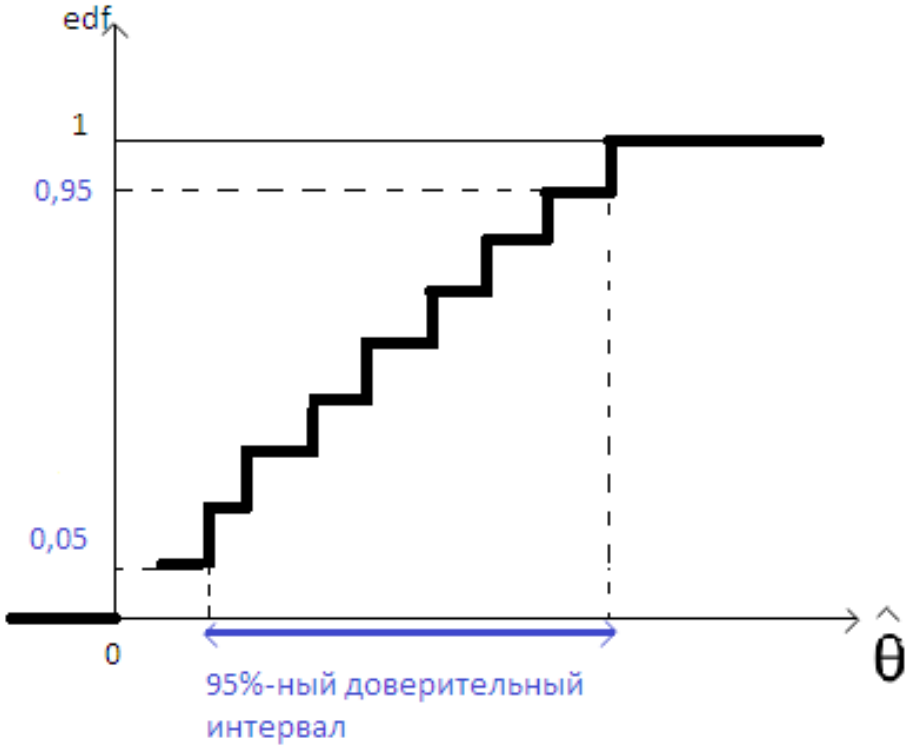
$$\hat{\theta}^* = \begin{cases} 0 \text{ с вероятностью } 1/4 \\ 0{,}8 \text{ с вероятностью } 1/2 \\ 1 \text{ с вероятностью } 1/4 \end{cases}$$

Простой пример (3/3)

График эмпирической функции распределения оценки:



Если «ступеней» больше:



- Если наблюдений n, то количество вариантов бутстраповской статистики n^n

- задача слишком сложная

- Выход - симуляции

Симуляции: пример с медианой

Допустим, мы хотим бутстрапировать некоторую статистику $\hat{\varphi} = \hat{\varphi}(\{x1, x2, \dots, x_n\})$.

Выберем B – количество будущих выборок.

Для каждого b=1,2,…,B построим бутстраповскую выборку $\{x_1^*, x_2^*, \dots, x_n^*\}$, вытягивая ее элементы случайным образом с возвращением из исходной выборки $\{x1, x2, \dots, x_n\}$.

Вычислим бутстраповскую статистику

$$\widehat{\varphi_b^*} = \hat{\varphi}^*(\{x_1^*, x_2^*, \dots, x_n^*\}_b).$$

Симуляции: доверительный интервал

- Для получения квантилей отсортировать бутстраповские статистики в порядке возрастания.

- В качестве квантилей $q^*_{\alpha/2}$ и $q^*_{1-\alpha/2}$ взять значения
- $\widehat{\varphi^*_{[B*\frac{\alpha}{2}]}}$ и $\widehat{\varphi}^*_{[B*\left(1-\frac{\alpha}{2}\right)+1]}$
- где [.] означает взятие целой части.

Итог: без какой-либо теории и дополнительных предпосылок получили доверительный интервал для оценки! Этот метод называется «дикий бутстрап».

Симуляции: каким должно быть B?

- Правило «большого пальца» (Кэмерон, Триведи, гл. 11):

$$B = \left(\frac{1{,}96}{0{,}1}\right)^2 * \frac{2+\gamma}{4}$$

- Обеспечивает относительное отклонение бутстраповской статистики от статистики, посчитанной при бесконечном числе повторений, меньше чем на 10% в вероятностью не менее 95%.

γ - коэффициент эксцесса – мера островершинности распределения относительно нормального распределения:

$$\gamma = \frac{E(x - E(x))^4}{\sigma^4} - 3$$

B=384*(2+0)/4=960 для нормального распределения.

Автоматические значения в пакетах 1000, 3000, 5000, 10000.

Симуляции: каким должно быть B?

- *B* должно быть как можно больше. Нельзя брать 100, 200…
- И всё-таки есть Теорема ☺: при выполнении «некоторых условий» разница между confidence level номинальным (95%) и confidence level по процедуре (не факт, что 95%) пропорциональна $1/\sqrt{n}$.
- Т.е. если хочу снизить эту разницу в 10 раз, то выборку надо увеличить в 100 раз.
- Доказывать не будем

Доверительные интервалы

Эфронов доверительный интервал («дикий», «наивный» бутстрап)

бутстрапируем	саму оценку
вытягиваем	x^*
считаем	$\hat{\theta}_b^*$
повторяем	В раз
строим распределение для	$\left\{\hat{\theta}_b^*\right\}_{b=1}^{B}$
Интервал	$\theta \in \left[q_{\alpha/2}; q_{1-\alpha/2}\right]$

Доверительный интервал Холла

бутстрапируем	Отклонение оценки от истинного значения
вытягиваем	x^*
считаем	$\hat{\theta}_b^* - \hat{\theta}$
повторяем	B раз
строим распределение для	$\left\{\hat{\theta}_b^* - \hat{\theta}\right\}_{b=1}^{B}$
Интервал	$\theta \in \left\lfloor \hat{\theta} - q_{1-\alpha/2}; \hat{\theta} - q_{\alpha/2} \right\rfloor$

t-процентильный доверительный интервал

Бутстрапируем t-статистистику	$\frac{\hat{\theta}-\theta}{s.e.(\hat{\theta})}$
вытягиваем	x^*
Считаем	$\frac{\hat{\theta}_b^*-\hat{\theta}}{s.e.(\hat{\theta}_b^*)}$
строим распределение для бутстрап-аналога t-статистики	$\left\{\frac{\hat{\theta}_b^*-\hat{\theta}}{s.e.(\hat{\theta}_b^*)}\right\}_{b=1}^{B}$
Интервал	$\theta \in \left[\hat{\theta}-s.e.(\hat{\theta})*q_{1-\alpha/2};\hat{\theta}-s.e.(\hat{\theta})*q_{\alpha/2}\right]$

t-процентильный доверительный интервал

- Пример на доске – доверительный интервал. P(Yi>0).
- Есть Теорема: при выполнении «некоторых условий» разница между confidence level номинальным (95%) и confidence level по процедуре (не факт, что 95%) пропорциональна 1/n.
- Т.е. если хочу снизить эту разницу в 10 раз, то выборку надо увеличить в 10 раз.
- Доказывать не будем
- (+) Выборка нужна меньше, чем в диком бутстрапе
- (-) Откуда брать s.e. оценки, если её формула неизвестна?

Симметричный t-процентильный д. и. (подходит для тестирования гипотез)

бутстрапируем	$\left\lvert \frac{\hat{\theta}-\theta}{s.e.(\hat{\theta})} \right\rvert$
вытягиваем	x^*
Считаем	$\frac{\hat{\theta}_b^* - \hat{\theta}}{s.e.(\hat{\theta}_b^*)}$
строим распределение для	$\left\{ \frac{\hat{\theta}_b^* - \hat{\theta}}{s.e.(\hat{\theta}_b^*)} \right\}_{b=1}^{B}$
Интервал	$\theta \in \left\lfloor \hat{\theta} - s.e.(\hat{\theta}) * q_{1-\alpha}; \hat{\theta} + s.e.(\hat{\theta}) * q_{1-\alpha} \right\rceil$

Чем закончилась история с порогом долга?

- В апреле 2013 г. Herndon, Ash, Pollin нашли ошибку в расчётах K. Reinhart & Kenneth S. Rogoff «Growth in a time of debt» (2010) и опровергли основной результат их исследования. На основании тех же самых данных они получили, что влияние государственного долга на темпы роста реального ВВП отрицательное и одинаково для любых значений долга (монотонная зависимость).
- Таким образом, был поставлен под сомнение один из аргументов в пользу необходимости «политики затягивания поясов» в Европе, которым выступало исследование R&R.
- Но эта статья 2013г. не отменяет результатов многочисленных исследований, нашедших нелинейную зависимость для разных групп стран.

```
#bootstrap руками делаем цикл
B<-1000 #число бутстраповских выборок
n<-NROW(mydata)  #размер бутстраповской выборки совпадает с размером исходной выборки
tip<-NULL #цикл нужно с чего-то начать, создаём пустой вектор


for (i in 1:B){
   bootID<-sample(c(1:n),n,replace = TRUE)        # случайная выборка с возвращением
    tip[i] = (-lm(mydata[bootID,1] ~ mydata[bootID,2] +
    I((mydata[bootID,2])^2))$coef[2])/(2*lm(mydata[bootID,1] ~
    mydata[bootID,2] + I((mydata[bootID,2])^2))$coef[3])
    #для каждого i оценили модель и посчитали отношение оценок коэффициентов}
#95% confidence interval
tip_sorted<-sort(tip)
lower_bound<-tip_sorted[round(B*0.025)]
upper_bound<-tip_sorted[round(B*0.975)]
```

Содержание

- Зачем всё это? Пример из жизни
- Решение проблемы 1: бутстрап
- Решение проблемы 2: Дельта-метод
- Что почитать?
- Чем закончилась история

Дельта-метод: теоретическое обоснование

- Центральная предельная теорема

Let $\{z_n\}$ *be IID with* $\mathbb{E}[z_i] = \mu$ *and* $\mathbb{V}[z_i] = \sigma^2$. *Then,*

$$\sqrt{n}\left(\frac{1}{n}\sum z_i - \mu\right) \xrightarrow{d} N(0, \sigma^2),$$

as $n \to \infty$.

Дельта-метод: теоретическое обоснование

- Теорема Слуцкого

If $z_n \xrightarrow{d} z$ *and* $c_n \xrightarrow{p} c$ *as* $n \to \infty$, *then*

1. $z_n + c_n \xrightarrow{d} z + c$
2. $z_n c_n \xrightarrow{d} zc$
3. $\frac{z_n}{c_n} \xrightarrow{d} \frac{z}{c}$ *if* $c \neq 0$

Дельта-метод: теоретическое обоснование

- Ряд Тейлора
- Функция g(x) непрерывна и дважды дифференцируема для любого x из X
- Для x0 имеем:

$$g(x) = g(x_0) + g'(x_0)(x - x_0) + \frac{1}{2!}g''(x_0)(x - x_0)^2 + o(x^2)$$

Одномерный дельта-метод

Let $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \xi$. What is the asymptotic distribution of $g(\hat{\mu})$?

1. Apply Taylor Expansion at μ

$$g(\hat{\mu}) = g(\mu) + g'(\mu)(\mu - \mu) + o(\hat{\mu})$$

2. Re-arrange the terms

$$g(\hat{\mu}) - g(\mu) = g'(\mu)(\mu - \mu) + o(\hat{\mu})$$

$$\sqrt{n}\left(g(\hat{\mu}) - g(\mu)\right) = g'(\mu)\sqrt{n}(\mu - \mu) + \sqrt{n}o(\hat{\mu})$$

3. Use $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \xi$ Then,

$$\sqrt{n}\left(g(\hat{\mu}) - g(\mu)\right) \xrightarrow{d} g'(\mu)\xi$$

Assume $\xi \sim N(0, \sigma^2)$. Then,

$$\sqrt{n}\left(g(\hat{\mu}) - g(\mu)\right) \xrightarrow{d} N(0, \left(g'(\mu)\right)^2 \sigma^2)$$

https://www.econ.msu.ru/ext/lib/Category/x81/xf9/33273/file/%D0%92%D1%81%D1%82%D1%80%D0%B5%D1%87%D0%B0%202_%20DeltaMethod.pdf Алексей Хазанов

Многомерный дельта-метод

If $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \xi$, *where* $g(u)$ *is continuously differentiable in a neighborhood of* μ *then as* $n \to \infty$

$$\sqrt{n}\left(g(\hat{\mu}) - g(\mu)\right) \xrightarrow{d} G'\xi,$$

where $G(u) = \frac{\partial}{\partial u} g(u)'$ *and* $G = G(\mu)$. *In particular, if* $\xi \sim N(0, V)$, *then as* $n \to \infty$

$$\sqrt{n}\left(g(\hat{\mu}) - g(\mu)\right) \xrightarrow{d} N(0, G'VG)$$

https://www.econ.msu.ru/ext/lib/Category/x81/xf9/33273/file/%D0%92%D1%81%D1%82%D1%80%D0%B5%D1%87%D0%B0%202_%20DeltaMethod.pdf Алексей Хазанов

Дельта-метода: реализация в R и проблема

- Реализация в R:

пакет car, команда deltaMethod

- Проблема:

А что делать, если выборка мала и асимптотика неприменима?

- Бутстрап

Что почитать или посмотреть

- С. Анатольев ликбез + список базовой литературы по бутстрапу в журнале «Квантиль»: http://quantile.ru/03/03-SA.pdf
- Б.Б. Демешев Лекция «Наивный бутстрап» https://youtu.be/wIPq_OoYcjc
- Дельта-метод популярно: https://www.statlect.com/asymptotic-theory/delta-method
- А. А. Хазанов Лекция «Дельта-метод» https://www.econ.msu.ru/ext/lib/Category/x81/xf9/33273/file/%D0%92%D1%81%D1%82%D1%80%D0%B5%D1%87%D0%B0%202_%20DeltaMethod.pdf
- Пакет boot для R https://cran.r-project.org/web/packages/boot/index.html
- Пакет для дельта-метода в R: https://cran.r-project.org/web/packages/modmarg/vignettes/delta-method.html