Эконометрический ликбез: бутстрап

Основы бутстрапирования[⋆]

Станислав Анатольев[†]

Российская экономическая школа, Москва, Россия

Настоящее эссе – введение в принципы и методологию бутстрапа. Даются основы бутстраповской инференции, ресэмплинга, асимптотического рафинирования. Изложение сопровождается поясняющими примерами. Приводятся краткий анонс методологических эссе данного выпуска журнала «Квантиль» и ссылки на незатронутый материал.

1 Приближение бутстрапом

Не секрет, что точная инференция[1] в современном эконометрическом анализе практически недостижима, если модель хоть сколь-нибудь замысловатая, а вовлеченные распределения не являются нормальными. Среди приближенных, конечно, лидирует асимптотический подход, т.е. когда при инференции используются приближения, свойственные теории больших выборок. Асимптотическая теория, правда, частенько тоже разочаровывает неточностью предоставляемых приближений или трудностью аналитического вывода асимптотических результатов (см. Анатольев, 2005). В этих случаях используют метод *бутстрапирования*, предоставляющий альтернативные асимптотическим приближения или возможность обходиться без сложных аналитических выводов.

Формально изобретение бутстрапа приписывают Эфрону (Efron, 1979, см. также Эфрон, 1988), хотя и ранее можно было встретить интуитивные применения бутстраповских идей. Например, в Cowles (1934) автор использовал колоду карт для генерации случайной подвыборки активов для последующего сравнения ее доходности с доходностью подвыборки, построенной по определенным правилам. Кстати, интенсивное использование при бутстрапировании симуляций (экспериментов Монте-Карло) является той причиной, по которой многие прикладные исследователи считают бутстрап методом Монте-Карло. На самом деле суть бутстрапа вовсе не в симуляциях.

В основе бутстраповского подхода лежит та идея, что истинное распределение данных можно хорошо приблизить эмпирическим, т.е. тем, как данные легли в выборке. На самом деле, эмпирическое распределение – единственный источник информации об истинном распределении данных, что у исследоваетеля есть помимо модели. Идеологически бутстрап как раз и подразумевает это приближение истинного распределения эмпирическим; техническая же сторона дела – трансформировать приближение для распределения данных в приближенное распределение интересующих исследователя статистик, где симуляции и используются.

Идея того, что распределение самих данные используются вместо их истинного распределения нашла отражение в названии метода. «Бутстрапами» называются ремешки на обуви, ухватясь за которые, барон Мюнхаузен, по слухам, вытащил себя из болота. Не очень ясно,

---

[⋆]Работа основана на лекциях, читаемых автором в РЭШ. Автор благодарит Александра Цыплакова за полезные замечания. Цитировать как: Анатольев, Станислав (2007) «Основы бутстрапирования», Квантиль, №3, стр. 1–12. Citation: Anatolyev, Stanislav (2007) “The basics of bootstrapping,” Quantile, No.3, pp. 1–12.

[†]Адрес: 117418, г. Москва, Нахимовский проспект, 47, офис 1721(3). Электронная почта: *sanatoly@nes.ru*

[1]Инференция включает в себя построение доверительных областей и проверку статистических гипотез.

правда, не за волосы ли себя он тащил, и почему в электронной версии книги не присутствуют ни болото (кроме эпизода охоты на какого-то зверя в трясине), ни бутстрап вовсе.[2]

Итак, пусть из исходной популяции случайной величины z с распределением $F(z)$ получена выборка размера n. Тогда (кумулятивная) эмпирическая функция распределения (ЭФР)

$$F^*(z) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{[z_i \le z]},$$

где $\mathbb{I}_{[\cdot]}$ – индикатор-функция, равномерно почти наверное стремится к $F(z)$ при $n \to \infty$, что составляет суть леммы Гливенко–Кантелли. Это свойство мотивирует использование бутстрапа и его благоприятные свойства.

Чтобы более наглядно пояснить бутстраповский метод, рассмотрим простейший пример. Пусть в выборке всего два наблюдения:

$$(x_1, y_1) = (1, 2);\ (x_2, y_2) = (1, 2).$$

Допустим, нас интересует коэффициент θ линейной проекции y на x без свободного члена, т.е. $y_i = \theta x_i + \varepsilon_i$, $\mathbb{E}[x_i\varepsilon_i] = 0$. В этом случае МНК-оценка параметра θ равна

$$\widehat{\theta} = \frac{x_1y_1 + x_2y_2}{x_1^2 + x_2^2} = \frac{1\times 2 + 2\times 1}{1^2 + 2^2} = \frac{4}{5}.$$

Эмпирическая функция распределения данных представляет собой

$$(x, y) = \begin{cases} (1,2) & \text{с вероятностью } 1/2, \\ (2,1) & \text{с вероятностью } 1/2. \end{cases}$$

По отношению к этому распределению данные выборки из двух наблюдений распределены следующим образом:

$$(x_1^*, y_1^*);\ (x_2^*, y_2^*) = \begin{cases} (1,2);\ (1,2) & \text{с вероятностью } 1/4, \\ (2,1);\ (2,1) & \text{с вероятностью } 1/4, \\ (1,2);\ (2,1) & \text{с вероятностью } 1/4, \\ (2,1);\ (1,2) & \text{с вероятностью } 1/4. \end{cases}$$

Это распределение является бутстраповским распределением (бутстраповской) выборки. Соответственно, бутстраповская МНК-оценка распределена как

$$\widehat{\theta}^* = \begin{cases} 1/2 & \text{с вероятностью } 1/4, \\ 4/5 & \text{с вероятностью } 1/2, \\ 2 & \text{с вероятностью } 1/4. \end{cases}$$

Это и есть бутстраповское распределение $\widehat{\theta}$.

2 Приближение симуляциями

Пример, рассмотренный нами в предыдущем разделе, был чрезвычайно прост: размер исходной выборки был равен 2. В общем случае, когда в выборке n наблюдений, количество вариантов значений бутстраповской статистки имеет порядок n^n. Таким образом, в вычислительном плане задача сильно усложняется по мере роста n. К решению этой задачи можно привлечь компьютер, но не стоит ему поручать эту задачу в буквальном смысле. Даже если

[2] Rudolph Erich Raspe "The Surprising Adventures of Baron Munchausen." Электронная книга от Authorama: *http://www.authorama.com/book/adventures-of-baron-munchausen.html.*

компьютер справится с этой сложной комбинаторной задачей за разумное время, полученный результат окажется для наших целей точным сверх необходимого. Гораздо выгодней воспользоваться дополнительным приближением, на этот раз с помощью симуляций. Эта идея прекрасна тем, что как раз из эмпирического распределения, присвающего всем наблюдениям выборки равные веса, вытягивать «наблюдения» (*ресэмплировать*) просто и удобно.

Предположим, мы хотим пробутстрапировать статистику $\widehat{\varphi} = \widehat{\varphi}(\{z_1; \ldots; z_n\})$. Выберем количество будущих бутстраповских выборок B (обычно хватает порядка 1000). Для каждого $b = 1, 2, \ldots, B$ построим *бутстраповскую выборку* $\{z_1^*; z_2^*; \ldots; z_n^*\}_b$, вытягивая ее элементы случайным образом с возвращением из исходной выборки $\{z_1; \ldots; z_n\}$, и вычислим *бутстраповскую статистику* $\widehat{\varphi}_b^* = \widehat{\varphi}(\{z_1^*; \ldots; z_n^*\}_b)$. Полученный набор бутстраповских статистик $\widehat{\varphi}_1^*, \ldots, \widehat{\varphi}_B^*$ с приписанием каждой веса $1/B$ составляет (приближенное) бутстраповское распределение статистики $\widehat{\varphi}$. Чаще всего бутстраповское распределение используют для получения бутстраповских квантилей, для чего нужно лишь отсортировать бутстраповские статистики в порядке возрастания и в качестве квантилей $q_{\alpha_1}^*$, $q_{1-\alpha_2}^*$ взять значения $\widehat{\varphi}_{\lfloor B\alpha_1 \rfloor}^*$, $\widehat{\varphi}_{\lfloor B(1-\alpha_2)+1 \rfloor}^*$, где $\lfloor \cdot \rfloor$ означает взятие целой части.

Сгенерировать дискретную случайную величину, равномерно распределенную на множестве индексов от 1 до n просто: надо сначала сгенерировать случайную величину, распределенную равномерно на $[0, 1]$, затем присвоить значение индекса, соответствующее попаданию в отрезки $[0, 1/n]$, $[1/n, 2/n], \cdots$, $[1 - 1/n, 1]$. Например, в программе GAUSS запись `Z[ceil(n*rndu(n,1)),.]` возвращает готовую матрицу с бутстраповский выборкой из строк матрицы `Z`.

Может возникнуть вопрос – а почему бы предварительно не сгладить «лесенку» ЭФР прежде чем ресэмплировать. Предложение сомнительное, ибо если полезность этого шага неочевидна, то процесс ресэмплинга технически сильно усложняется.

3 Рецентрирование

Вернемся к идеологии бутстрапа. При его реализации существуют некоторые тонкие моменты, игнорируя которые, можно легко все испортить. Один из таких тонких моментов – необходимость *рецентрирования*, когда речь идет о разностях (или расстояниях) между выборочными и популяционными объектами. Предположим, мы бутстрапируем разность $\widehat{\theta} - \theta$, где $\widehat{\theta}$ – аналоговая[3] оценка параметра θ. При бутстрапировании $\widehat{\theta}$ мы создаем бутстраповские статистики, подсчитываемые по той же формуле, что и $\widehat{\theta}$, но на бутстраповских выборках. А вот бутстраповским аналогом параметра θ является не он сам, а первоначальная статистика $\widehat{\theta}$. Действительно, если при приближении бутстрапом истинное распределение искривляется в эмпирическое, то и истинный параметр искривляется в свою оценку. Итак, правильным бутстраповским аналогом разности $\widehat{\theta} - \theta$ является, таким образом, $\widehat{\theta}^* - \widehat{\theta}$, но ни в коем случае не $\widehat{\theta}^* - \theta$. Это и есть рецентрирование.

Даже имеющий верное представление о рецентрировании исследователь может ошибиться, действуя механически. Например, бутстрапируя обычную t-статистику

$$t = \frac{\widehat{\theta}}{\mathrm{se}(\widehat{\theta})},$$

легко забыть о наличии в числителе разницы между $\widehat{\theta}$ и нулем, истинным значением параметра θ при нулевой гипотезе

$$H_0 : \theta = 0.$$

[3] Пусть интересующий нас параметр θ известным образом зависит от функции распределения z, $F(z)$. Согласно принципу аналогий, аналоговая оценка $\widehat{\theta}$ строится с помощью замены истинной функции распределения $F(z)$ на ее эмпирический аналог $F^*(z)$. Например, если нас интересует популяционное среднее $\mathbb{E}[z] = \int_{-\infty}^{+\infty} z dF(z)$, то его аналоговая оценка будет равна $\widehat{\mathbb{E}[z]} = \int_{-\infty}^{+\infty} z dF_n(z) = n^{-1} \sum_{i=1}^{n} z_i$.

Правильная версия бутстраповского аналога t-статистики есть, таким образом,

$$t^* = \frac{\widehat{\theta}^* - \widehat{\theta}}{\mathrm{se}^*(\widehat{\theta})}.$$

Причем точно так же выглядит бутстраповская t-статистика для гипотезы $H_0 : \theta = 13$, например, хотя сама t-статистика, разумеется, выглядит несколько по-другому.

Приведем другие, более сложные примеры рецентрирования, не акцентируясь на обозначениях, о значении которых читатель сам догадается. Положим, нас интересует нулевая гипотеза

$$H_0 : g(\theta) = 0.$$

Бутстраповский аналог статистики Вальда

$$\mathcal{W} = ng(\widehat{\theta})' \left(\widehat{G}\widehat{V}_\theta\widehat{G}'\right)^{-1} g(\widehat{\theta})$$

есть

$$\mathcal{W}^* = n\left(g(\widehat{\theta}^*) - g(\widehat{\theta})\right)' \left(\widehat{G}^*\widehat{V}_\theta^*\widehat{G}^{*\prime}\right)^{-1} \left(g(\widehat{\theta}^*) - g(\widehat{\theta})\right).$$

Бутстраповский аналог статистики отношения правдоподобия

$$\mathcal{LR} = 2\left(\max_\theta \ell_n(\theta) - \max_{\theta\colon g(\theta)=0} \ell_n(\theta)\right)$$

есть

$$\mathcal{LR}^* = 2\left(\max_\theta \ell_n^*(\theta) - \max_{\theta\colon g(\theta)=g(\widehat{\theta})} \ell_n^*(\theta)\right).$$

Бутстраповский аналог статистики разницы оптимумов, основанной на ОММ,

$$\mathcal{DD} = n\left[\min_{\theta\colon g(\theta)=0} \mathcal{Q}_n(\theta) - \min_\theta \mathcal{Q}_n(\theta)\right],$$

где $\mathcal{Q}_n(\theta)$ – целевая функция ОММ:

$$\mathcal{Q}_n(\theta) = \left(\frac{1}{n}\sum_{i=1}^n m\left(z_i, \theta\right)\right)' \widehat{\Sigma}^{-1} \left(\frac{1}{n}\sum_{i=1}^n m\left(z_i, \theta\right)\right),$$

и где $\widehat{\Sigma}$ состоятельно оценивает $\Sigma = \mathbb{E}\left[m\left(z, \theta\right) m\left(z, \theta\right)'\right]$, есть

$$\mathcal{DD}^* = n\left[\min_{\theta\colon g(\theta)=g(\widehat{\theta})} \mathcal{Q}_n^*(\theta) - \min_\theta \mathcal{Q}_n^*(\theta)\right],$$

где $\mathcal{Q}_n^*(\theta)$ – бутстраповская целевая функция ОММ:

$$\mathcal{Q}_n^*(\theta) = \left(\frac{1}{n}\sum_{i=1}^n m\left(z_i^*, \theta\right) - \frac{1}{n}\sum_{i=1}^n m(z_i, \widehat{\theta})\right)' \widehat{\Sigma}^{*-1} \left(\frac{1}{n}\sum_{i=1}^n m\left(z_i^*, \theta\right) - \frac{1}{n}\sum_{i=1}^n m(z_i, \widehat{\theta})\right).$$

Заметим, что в последнем примере встретилось сразу два случая рецентрирования.

Отметим еще один момент. Может возникнуть вопрос, почему мы не рецентрируем, например, в знаменателе $\mathrm{se}^*(\widehat{\theta})$ бутстраповской t-статистики или в бутстраповских пивотизирующих матрицах $\widehat{V}_\theta^*$ и $\widehat{\Sigma}^*$. Дело в том, что это непринципиально, ибо это множители, от которых всего лишь требуется сходимость по вероятности к неслучайным пределам. В случае же, когда речь идет именно о расстояниях (разностях), рецентрирование принципиально.

В следующих двух разделах мы познакомимся с примерами использования бутстрапа.

4 Бутстраповская корректировка смещения

Сначала рассмотрим нетипичное применение бутстрапа, не связанное с инференцией, но помогающее еще глубже понять принципы его работы и использования. Бутстрап позволяет скорректировать смещение, связанное с конечностью выборки. Пусть у нас есть состоятельная оценка $\widehat{\theta}$ параметра θ, построенная согласно принципу аналогий, причем $\mathbb{E}[\widehat{\theta}] \neq \theta$. Смещение оценки $\widehat{\theta}$ равно $\mathbb{B}[\widehat{\theta}] = \mathbb{E}[\widehat{\theta}] - \theta$. Бутстраповским аналогом этого смещения является величина $\mathbb{B}^*[\widehat{\theta}] = \mathbb{E}^*[\widehat{\theta}^*] - \widehat{\theta}$ (вновь видим рецентрирование!), где $\mathbb{E}^*$ обозначает математическое ожидание по отношению к ЭФР. Если это бутстраповское смещение вычислить, то можно будет скорректировать исходную статистику на смещение:

$$\widehat{\theta}_{BC} = \widehat{\theta} - \mathbb{B}^*[\widehat{\theta}] = 2\widehat{\theta} - \mathbb{E}^*[\widehat{\theta}^*].$$

Рассмотрим пару примеров.

Пусть случайная величина z имеет среднее μ. Исследователь оценивает μ с помощью $\bar{z}_n$, а μ^2 – с помощью $\bar{z}_n^2$. Как выглядит бутстраповская корректировка смещения этих двух оценок? В первом случае, $\bar{z}_n^*$, бутстраповский аналог $\bar{z}_n$, имеет среднее $\bar{z}_n$ по отношению к эмпирическому распределению: $\mathbb{E}^*\left[\bar{z}_n^*\right] = \bar{z}_n$. Поэтому бутстраповский аналог смещения (которое само по себе – ноль) есть $\mathbb{B}^*[\bar{z}_n] = \mathbb{E}^*\left[\bar{z}_n^*\right] - \bar{z}_n = 0$, так что оценка μ, скорректированная на смещение, есть $\bar{z}_n - \mathbb{B}^*[\bar{z}_n] = \bar{z}_n$, что логично. Во втором случае, смещение $\bar{z}_n^2$ равно

$$\mathbb{B}[\bar{z}_n^2] = \mathbb{E}\left[\bar{z}_n^2\right] - \mu^2 = \mathbb{V}\left[\bar{z}_n\right] = \frac{1}{n}\mathbb{V}\left[z\right],$$

поэтому бутстраповский аналог смещения – выборочный аналог этой величины:

$$\mathbb{B}^*[\bar{z}_n^2] = \frac{1}{n}\mathbb{V}^*\left[z\right] = \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^n z_i^2 - \bar{z}_n^2\right),$$

так что оценкой μ^2, скорректированной на смещение, будет

$$\bar{z}_n^2 - \mathbb{B}^*[\bar{z}_n^2] = \frac{n+1}{n}\bar{z}_n^2 - \frac{1}{n^2}\sum_{i=1}^n z_i^2.$$

В этих простых примерах удалось аналитически вывести бутстраповское смещение. В более типичных ситуациях это сделать невозможно, и на практике прибегают к приближению симуляциями, оценивая $\mathbb{E}^*[\widehat{\theta}^*]$ с помощью усреднения по бутстраповским выборкам:

$$\widehat{\theta}_{BC} = 2\widehat{\theta} - \frac{1}{B}\sum_{b=1}^B \widehat{\theta}_b^*.$$

Еще раз подчеркнем, что бутстрап способен удалить существенную часть смещения, вызванную конечностью выборки, но не способен справиться с асимптотическим смещением. Так что если оценка несостоятельна, бутстрап бессилен.

5 Бутстраповская инференция

Основное использование бутстрапа – это, конечно же, инференция. При этом, по сути, вся бутстраповская технология используется лишь для получения одного или двух бутстраповских критических значений, используя которые наряду с первоначальной оценкой и, возможно, стандартными ошибками или пивотизирующей матрицей, строят бутстраповские доверительные интервалы или проверяют гипотезы.

Рассмотрим несколько вариантов бутстраповских статистик, используемых для построения доверительных интервалов, и подчеркнем их положительные и отрицательные качества. Пусть нас интересует построение статистических выводов относительно параметра θ на основе ее состоятельной оценки $\widehat{\theta}$.

Пробутстрапируем статистику $\widehat{\theta}-\theta$, бутстраповский аналог которой $\widehat{\theta}^*-\widehat{\theta}$. Получим бутстраповские квантили $q^{*\%}_{\alpha_1}$ и $q^{*\%}_{1-\alpha_2}$ и построим бутстраповский *процентильный* доверительный интервал с покрытием $1-\alpha_1-\alpha_2$

$$CI_{\%}=\left[\widehat{\theta}-q^{*\%}_{1-\alpha_2},\ \widehat{\theta}-q^{*\%}_{\alpha_1}\right].$$

Заметим, что неправильно бутстрапировать нецентрированную статистику $\widehat{\theta}$, ибо использующий квантили ее бутстраповского аналога $\widehat{\theta}^*$ так называемый *эфронов* доверительный интервал плох (кроме случая, характеризуемого симметричным распределением), ибо, не используя рецентрирование, только усиливает смещение, присущее исходной выборке.

Обратим внимание, что процентильный доверительный интервал не требует подсчета стандартных ошибок, что удобно, особенно в случаях, когда их трудно вывести или подсчитать. Тем не менее, есть причины (их мы обсудим позже), по которым стоит бутстрапировать t-статистики, если стандартные ошибки возможно качественно подсчитать. Итак, пробутстрапируем статистику $(\widehat{\theta}-\theta)/\mathrm{se}(\widehat{\theta})$, бутстраповский аналог которой $(\widehat{\theta}^*-\widehat{\theta})/\mathrm{se}^*(\widehat{\theta})$. Получим бутстраповские квантили $q^{*\%t}_{\alpha_1}$ и $q^{*\%t}_{1-\alpha_2}$ и построим бутстраповский *t-процентильный* доверительный интервал

$$CI_{\%t}=\left[\widehat{\theta}-\mathrm{se}(\widehat{\theta})q^{*\%t}_{1-\alpha_2},\ \widehat{\theta}-\mathrm{se}(\widehat{\theta})q^{*\%t}_{\alpha_1}\right].$$

Далее, чаще всего в приложениях $\alpha_1=\alpha_2=\alpha/2$. В таком случае можно бутстрапировать не центрированную оценку или t-статистику, а их модули. Например, пробутстрапировав статистику $|\widehat{\theta}-\theta|$, бутстраповский аналог которой равен $|\widehat{\theta}^*-\widehat{\theta}|$, получим бутстраповскую квантиль $q^{*|\%|}_{1-\alpha}$ и построим бутстраповский *симметричный процентильный* доверительный интервал с покрытием $1-\alpha$

$$CI_{|\%|}=\left[\widehat{\theta}-q^{*|\%|}_{1-\alpha},\ \widehat{\theta}+q^{*|\%|}_{1-\alpha}\right].$$

Аналогично, пробутстрапировав статистику $|\widehat{\theta}-\theta|/\mathrm{se}(\widehat{\theta})$, бутстраповский аналог которой $|\widehat{\theta}^*-\widehat{\theta}|/\mathrm{se}^*(\widehat{\theta})$, получим бутстраповскую квантиль $q^{*|\%t|}_{1-\alpha}$ и построим бутстраповский *симметричный t-процентильный* доверительный интервал

$$CI_{|\%t|}=\left[\widehat{\theta}-\mathrm{se}(\widehat{\theta})q^{*|\%t|}_{1-\alpha},\ \widehat{\theta}+\mathrm{se}(\widehat{\theta})q^{*|\%t|}_{1-\alpha}\right].$$

В следующем разделе мы узнаем о причинах, по которым лучше строить симметричные версии доверительных интервалов (разумеется, если выполнено $\alpha_1=\alpha_2$), а также почему надо стремиться к бутстрапированию нормализованных разностей (t-статистик), а не просто разностей.

Теперь обсудим тестирование статистических гипотез при помощи бутстрапа. Пусть параметр скалярен, а нулевая гипотеза имеет простейший вид

$$H_0:\theta=\theta^0,$$

где θ^0 – конкретная величина. Если альтернативная гипотеза двусторонняя, $H_A:\theta\neq\theta^0$, то, пробутстрапировав модуль t-статистики $|\widehat{\theta}-\theta|/\mathrm{se}(\widehat{\theta})$, получим бутстраповское распределение $|\widehat{\theta}^*-\widehat{\theta}|/\mathrm{se}^*(\widehat{\theta})$ и ее квантиль $q^{*|\%t|}_{1-\alpha}$. Гипотеза H_0 отвергается, если $|\widehat{\theta}-\theta^0|/\mathrm{se}(\widehat{\theta})>q^{*|\%t|}_{1-\alpha}$. Если же альтернативная гипотеза односторонняя, например правосторонняя $H_A:\theta>\theta^0$, то приходится бутстрапировать t-процентильную статистику $(\widehat{\theta}-\theta^0)/\mathrm{se}(\widehat{\theta})$. Получив бутстраповское распределение $(\widehat{\theta}^*-\widehat{\theta})/\mathrm{se}^*(\widehat{\theta})$ и соответствующую квантиль $q^{*\%t}_{1-\alpha}$, отвергаем гипотезу H_0, если $(\widehat{\theta}-\theta^0)/\mathrm{se}(\widehat{\theta})>q^{*\%t}_{1-\alpha}$.

Если нулевая гипотеза имеет вид системы нелинейных ограничений $H_0: g(\theta)=0$, бутстрапируется статистика Вальда (см. раздел 3), и ее бутстраповская квантиль $q^{*\%W}_{1-\alpha}$ используется для принятия решения: если $\mathcal{W}>q^{*\%W}_{1-\alpha}$, гипотеза H_0 отвергается. Например, если система ограничений линейна,

$$H_0: R\theta=r,$$

то бутстраповская статистика Вальда принимает вид $(\widehat{\theta}^*-\widehat{\theta})'R'(R\widehat{V}^*_\theta R')^{-1}R(\widehat{\theta}^*-\widehat{\theta})$, и гипотеза отвергается, если $(R\widehat{\theta}-r)'(R\widehat{V}_\theta R')^{-1}(R\widehat{\theta}-r)>q^{*\%W}_{1-\alpha}$.

Заметим, что при асимптотическом подходе и доверительные интервалы, и решающие правила при тестировании гипотез выглядят точно так же, как и при бутстрапировании, за единственным исключением – используются другие квантили. Таким образом, вся процедура бутстрапирования, включая симуляции, ресэмплинг, подсчет бутстраповских статистик и т.д., подчинена единственной цели – получить правильные бутстраповские квантили.

6 Асимптотическое рафинирование

Говорят, что с помощью бутстрапа можно достичь асимптотического рафинирования. В этом разделе мы обсудим, что такое асимптотическое рафинирование и в каких случаях оно имеет место.

Пусть у нас есть некоторая статистика $\widehat{\varphi}$, истинное распределение которой $F_{\widehat{\varphi}}(z)$. Обозначим бутстраповское распределение этой статистики через $F^*_{\widehat{\varphi}}(z)$. Говорят, что с помощью бутстрапа достигается асимптотическое рафинирование, если ошибка аппроксимации истинного распределения $F_{\widehat{\varphi}}(z)$ бутстраповским $F^*_{\widehat{\varphi}}(z)$ большего порядка малости, чем ошибка аппроксимации асимптотическим распределением при стремлении объема выборки к бесконечности.

Инструментарием здесь являются так называемое *разложение Эджворта* (кумулятивной) функции распределения статистики вокруг ее предельного распределения по степеням $1/\sqrt{n}$. Рассмотрим асимптотически пивотальную t-статистику, т.е.

$$\widehat{\varphi}=\frac{\widehat{\theta}-\theta}{\mathrm{se}(\widehat{\theta})}\xrightarrow{d}N(0,1).$$

Для (кумулятивной) функции стандартного нормального распределения используем обычное обозначение $\Phi(z)$. Разложения Эджворта истинного и бутстраповского распределений вокруг асимптотического выглядят следующим образом:

$$F_{\widehat{\varphi}}(z)=\Phi(z)+\frac{h_1(z,F)}{\sqrt{n}}+\frac{h_2(z,F)}{n}+O\left(\frac{1}{n\sqrt{n}}\right),$$

$$F^*_{\widehat{\varphi}}(z)=\Phi(z)+\frac{h_1(z,F^*)}{\sqrt{n}}+\frac{h_2(z,F^*)}{n}+O\left(\frac{1}{n\sqrt{n}}\right).$$

Здесь $h_1(z,F)$ – четная по z, непрерывная по F функция, $h_2(z,F)$ – нечетная по z, непрерывная по F функция. Ошибки аппроксимации точного распределения асимптотическим и бутстраповским, соответственно, равны

$$\Phi(z)-F_{\widehat{\varphi}}(z)=\frac{h_1(z,F)}{\sqrt{n}}+O\left(\frac{1}{\sqrt{n}}\right)=O\left(\frac{1}{\sqrt{n}}\right),$$

$$F^*_{\widehat{\varphi}}(z) - F_{\widehat{\varphi}}(z) = \frac{h_1(z, F^*) - h_1(z, F)}{\sqrt{n}} + O\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right).$$

Здесь мы воспользовались тем фактом, что скорость сходимости разности $h_1(z, F^*) - h_1(z, F)$ есть $\sqrt{n}$ согласно уже упоминавшейся лемме Гливенко–Кантелли. Сравнивая порядки ошибок аппроксимации, делаем вывод, что в данном случае использование бутстрапа приводит к асимптотическому рафинированию. На практике это означает, что в достаточно больших выборках ошибка бутстраповского приближения, по идее, намного меньше, чем ошибка асимптотического приближения.

Рассмотрим теперь асимптотически непивотальную статистику

$$\widehat{\varphi} = \sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, V_\theta),$$

где непивотальность проявляется в зависимости предельного распределения от неизвестного параметра V_θ. Как и в предыдущем случае, разложим точное и бутстраповское распределения вокруг асимптотического:

$$F_{\widehat{\varphi}}(z) = \Phi(z, V_\theta) + \frac{h_1(z, F)}{\sqrt{n}} + O\left(\frac{1}{n}\right),$$

$$F^*_{\widehat{\varphi}}(z) = \Phi(z, V^*_\theta) + \frac{h_1(z, F^*)}{\sqrt{n}} + O\left(\frac{1}{n}\right).$$

Ошибки аппроксимации для асимптотического и бутстраповского распределений следующие:

$$\Phi(z, V_\theta) - F_{\widehat{\theta}}(z) = -\frac{h_1(z, F)}{\sqrt{n}} + O\left(\frac{1}{n}\right) = O\left(\frac{1}{\sqrt{n}}\right),$$

$$F^*_{\widehat{\theta}}(z) - F_{\widehat{\theta}}(z) = \Phi(z, V^*_{\widehat{\theta}}) - \Phi(z, V_{\widehat{\theta}}) + O\left(\frac{1}{n}\right) = O\left(\frac{1}{\sqrt{n}}\right),$$

ибо $V^*_{\widehat{\theta}} - V_{\widehat{\theta}}$ сходится к нулю со скоростью $\sqrt{n}$. Видим, что в данном случае использование бутстрапа не приводит к асимптотическому рафинированию. Итак, бутстрапирование асимптотически непивотальных статистик не дает асимптотического рафинирования.

А теперь рассмотрим модуль t-статистики

$$\widehat{\varphi} = \frac{|\widehat{\theta} - \theta|}{\mathrm{se}(\widehat{\theta})} \xrightarrow{d} |N(0, 1)|.$$

Вновь разложим точное и бутстраповское распределения по Эджворту:

$$F_{\widehat{\varphi}}(z) = 2\Phi(z) - 1 + \frac{2h_2(z, F)}{n} + O\left(\frac{1}{n\sqrt{n}}\right),$$

$$F^*_{\widehat{\varphi}}(z) = 2\Phi(z) - 1 + \frac{2h_2(z, F^*)}{n} + O\left(\frac{1}{n\sqrt{n}}\right).$$

Здесь ошибки аппроксимации для асимптотики и бутстрапа имеют порядки соответственно

$$2\Phi(z) - 1 - F_{\widehat{\theta}}(z) = O\left(\frac{1}{n}\right),$$

$$F^*_{\widehat{\theta}}(z) - F_{\widehat{\theta}}(z) = \frac{2}{n}\left(h_2(z, F^*) - h_2(z, F)\right) + O\left(\frac{1}{n\sqrt{n}}\right) = O\left(\frac{1}{n\sqrt{n}}\right).$$

Естественно, мы достигаем асимптотического рафинирования, но вдобавок заметим, что бутстрапирование симметричного двустороннего теста (модуля t-статистики) имеет ошибку большего порядка малости, чем бутстрапирование одностороннего (просто t-статистики).

Итак, теперь мы знаем причину, по которой желательно, если это возможно, пивотизировать и симметризовывать бутстрапируемые статистики.

7 Построение бутстраповских выборок (ресэмплинг) в регрессиях

Чаще всего исследователь производит инференцию в регрессионном контексте. Рассмотрим для начала кросс-секционную регрессию

$$y = x'\beta + e, \ E[e|x] = 0.$$

Пусть имеется набор пар $\{(x_i, y_i)\}_{i=1}^n$, представляющий из себя случайную выборку.

Самый непосредственный способ ресэмплинга, так называемый *непараметрический бутстрап*, – вытягивать n пар (x_i^*, y_i^*) из исходной выборки $\{(x_i, y_i)\}_{i=1}^n$. Ровно того же результата можно достичь ресэмплингом пар $(x_i^*, \widehat{e}_i^*)$ из набора $\{(x_i, \widehat{e}_i)\}_{i=1}^n$, где $\widehat{e}_i = y_i - x_i'\widehat{\beta}$ – МНК-остатки, и последующего восстановления переменных левой части с помощью модели: $y_i^* = x_i'^*\widehat{\beta} + \widehat{e}_i^*$. Нетрудно убедиться, что этот так называемый *остаточный бутстрап* численно идентичен непараметричекому. Зачем же он нужен? Дело в том, что в модели может присутствовать дополнительная информация в форме взаимоотношения регрессионных ошибок и регрессоров, и необходимо принять эту информацию во внимание при ресэмплинге, что удобно сделать именно через остаточный бутстрап. Например, если известно, что ошибки и регрессоры независимы, то эффективность бутстрапа можно увеличить, извлекая случайно с возвращением x_i^* из $\{x_i\}_{i=1}^n$ и $\widehat{e}_i^*$ из $\{\widehat{e}_i\}_{i=1}^n$ независимо друг от друга. А если вдобавок ошибки распределены нормально, т.е. $e_i \sim N(0, \sigma^2)$, то эффективность бутстрапа можно еще больше увеличить, извлекая бутстраповские ошибки не из остатков, а из нормального распределения $N(0, \widehat{\sigma}^2)$.

Теперь обсудим ресэмплинг при бутстрапе временных рядов. Временной ряд отличается от кросс-секционной выборки тем, что наблюдения (уравнения) здесь зависимы, и перемешивание при непараметрическом бутстрапе разрушает эту зависимость, так что вероятностная структура бутстраповских данных уже не имитирует вероятностную структуру исходных данных. Чтобы избежать этого, чаще всего используется *блочный бутстрап*, в котором бутстраповская выборка строится из блоков исходной выборки. Аналогично случаю независимых наблюдений, во временых рядах возможен остаточный бутстрап, однако такой способ применим только в тех редких случаях, когда регрессионные ошибки (инновации) серийно независимы.

Рассмотрим несколько альтернативных способов блочного ресэмплинга. Пусть $\{(x_t, y_t)\}_{t=1}^T$ – исходная выборка, а l – длина блока. Блоки могут быть перекрывающимися, тогда в первый блок войдут пары $(x_1, y_1), \ldots, (x_l, y_l)$, во второй – $(x_2, y_2), \ldots, (x_{l+1}, y_{l+1})$, в третий – $(x_3, y_3), \ldots, (x_{l+2}, y_{l+2})$, и наконец в $(T - l + 1)$-ый – наблюдения $(x_{T-l+1}, y_{T-l+1}), \ldots, (x_T, y_T)$. Блоки могут быть неперекрывающимися, тогда в первый блок войдут $(x_1, y_1), \ldots, (x_l, y_l)$, во второй – (x_{l+1}, y_{l+1}), $\ldots$, (x_{2l}, y_{2l}), и наконец, в последний $\lfloor \frac{T}{l} \rfloor$-ый блок – пары $y_{l\lfloor \frac{T}{l} \rfloor - l+1}, \ldots$, $y_{l\lfloor \frac{T}{l} \rfloor}$. При построении бутстраповской выборки блоки извлекаются случайно с возвращением, как отдельные пары извлекаются из выборки при непараметрическом бутстрапе в кросс-секциях. Длина блока выбирается исследователем и отражает компромисс между сохранением временной структуры ряда и разнообразием бутстраповских выборок.

Предыдущие два варианта блочного ресэмплинга нарушают стационарность ряда, т.е. из стационарной исходной выборки получаются нестационарные бутстраповские выборки из-за разрывов зависимости между блоками. Чтобы бутстраповские выборки получались стационарными, был предложен *стационарный бутстрап*, основанный на нефиксированной длине блоков. А именно, задается вероятность окончания блока p. Первый элемент бутстраповской выборки выбирается случайно; затем с вероятностью $1 - p$ в текущий блок включается следующий элемент исходной выборки, а с вероятностью p начинается новый блок, первый элемент которого снова выбирается случайно из исходной выборки. Так продолжается, пока в бутстраповскую выборку не будет набрано нужное количество элементов.

8 Что вошло в «Квантиль» и что осталось за кадром

Помимо данной работы, содержащей в основном вводный материал, настоящий выпуск журнала содержит еще три статьи по бутстрапу.

- Эссе Расселла Дэвидсона посвящено ресэмпинговым схемам в различных эконометрических моделях. Автор формулирует «золотые правила», которых следует придерживаться ради получения наиболее точной бутстраповской инференции. Несколько под другим углом обосновываются рецентрирование и пивотизация. Кроме того, автор обсуждает вычислительные аспекты, актуальные при проведении симуляций.
- Статья Питера Бюльмана обсуждает существующие ресэмпинговые схемы при бутстрапировании временных рядов: блочная, решетчатая и локальная бутстрап-схемы. Помимо теоретических свойств этих схем, автор приводит результаты экспериментов Монте-Карло, описывающие их поведение на практике.
- Наконец, работа Валентины Корради посвящена бутстрапу в контексте (обобщенного) метода моментов. Особое внимание уделяется временным рядам и асимптотическому рафинированию.

Некоторые специфические, хоть и важные, методы бутстрапирования недостаточно полно освещены в данном выпуске. Ниже приводятся ссылки на некоторые опубликованные работы.

- Бутстрапирование в моделях с порогами (Hansen, 1997) и структурными сдвигами (Hansen, 2000b) являются частными случаями общего подхода бутстрапа с фиксированными регрессорами для тестирования гипотезы линейности (Hansen, 2000a).
- Бутстрапирование условных моделей временных рядов включает в себя марковский бутстрап (Horowitz, 2003) и его упрощенную версию – бутстрапирование марковской цепью (Anatolyev & Vasnev, 2002).
- Бутстрапирование моделей VAR и функций импульсного отклика обсуждается в Kilian (1998) и Kilian (1999), а применение бутстрапа при прогнозировании – в Kim (1999).
- Бутстрапирование моделей с единичным корнем и соответствующих тестов обсуждается в Inoue & Kilian (2002) и Park (2003), а сеточный бутстрап для моделей с корнем, близким к единичному – в Hansen (1999).
- Использование бутстрапа для выбора модели из парных альтернатив продемонстрировано в Маккракен (2006).
- Бутстрапирование с целью борьбы с «добычей данных» разработано в White (2000), а применение к трейдинговым стратегиям – в Sullivan, Timmermann & White (1999).
- Бутстрапирование в моделях бинарного выбора исследуется в Wang, Wang & Carroll (1997), а в моделях с цензурированием – в Efron (1981).
- Различные варианты бутстрапирования моделей панельных данных (точнее, моделей компонент ошибки) описываются в Andersson & Karlsson (2001).
- Эффективное бутстрапирование в контексте условий на моменты, оцениваемых с помощью метода эмпирического правдоподобия или ОММ, привлекающее неравномерные веса при ресэмплинге, разработано в Brown & Newey (2002).

- Бутстрапирование непараметрических моделей и ядерных оценок обсуждается в Härdle & Bowman (1988).

- Бутстрапирование негладких функций статистик и применение, в частности, к медианной и квантильной регрессиям разработано в Hahn (1995) и Horowitz (1998).

- Близкому по духу к бутстрапу методу подвыборок посвящена книга Politis, Romano & Wolf (1999).

Наконец, существует довольно много обзорного материала по бутстрапу:

- Книги по бутстрапу в основном написаны статистиками: Efron & Tibshirani (1993), Davison & Hinkley (1997), Shao & Tu (1995).

- Эконометрическая специфика лучше отражена в обзорных статьях эконометристов, например, в Hall (1994) и Horowitz (2001).

- Отдельная монография Hall (1994) посвящена разложениям Эджворта. Хорошим пособием по подобным асимптотическим разложениям является учебник Барндорф-Нильсен & Кокс (1999).

- В сборнике статей Эфрон (1988) имеются переведенные на русский язык статьи Эфрона, включая классическую работу Efron (1979).

Список литературы

Анатольев, С.А. (2005). Асимптотические приближения в современной эконометрике. *Экономика и математические методы* 41, 84–94.

Барндорф-Нильсен О. & Д. Кокс (1999). *Асимптотические методы в математической статистике.* Москва: Мир.

Маккракен, М. (2006). Парные тесты на одинаковую точность прогнозов. *Квантиль* 1, 53–62.

Эфрон, Б. (1988). Нетрадиционные методы многомерного статистического анализа (сборник статей). Москва: Финансы и статистика.

Anatolyev, S. & A. Vasnev (2002). Markov chain approximation in bootstrapping autoregressions. *Economics Bulletin* 3, 1–8

Andersson, M. & S. Karlsson (2001). Bootstrapping error component models. *Computational Statistics* 16, 221–231.

Brown, B.W. & W.K. Newey (2002). Generalized method of moments, efficient bootstrapping, and improved inference. *Journal of Economic & Business Statistics* 20, 507–517.

Cowles, A. (1934). Can stock market forecasters forecast? *Econometrica* 1, 309–324.

Davison, A.C. & D.V. Hinkley (1997). *Bootstrap Methods and their Applications.* Cambridge University Press.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7, 1–26.

Efron, B. (1981). Censored data and the bootstrap. *Journal of American Statistical Association* 76, 312–319.

Efron, B. & R. Tibshirani (1993). *An Introduction to the Bootstrap.* Chapman and Hall.

Hahn, J. (1995). Bootstrapping quantile regression estimators. *Econometric Theory* 11, 105–121.

Hall, P. (1994). Methodology and theory for the bootstrap. Глава 39 в *Handbook of Econometrics* (под редакцией Engle, R. & D. McFadden), том 4. Elsevier Science.

Hansen, B.E. (1999). The grid bootstrap and the autoregressive model. *Review of Economics & Statistics* 81, 594–607.

Hansen, B.E. (2000a). Sample splitting and threshold estimation. *Econometrica* 68, 575–604.

Hansen, B.E. (2000b). Testing for structural change in conditional models. *Journal of Econometrics* 97, 93–115.

Hansen, B.E. (1997). Inference in TAR models. *Studies in Nonlinear Dynamics & Econometrics* 2, 1–14.

Härdle, W. & A. Bowman (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of American Statistical Assoication* 83, 102–110.

Horowitz, J.L. (2003). Bootstrap methods for Markov processes. *Econometrica* 71, 1049–1082.

Horowitz, J.L. (1998). Bootstrap methods for median regression models. *Econometrica* 66, 1327–1352.

Horowitz, J.L. (2001). The bootstrap. Глава 52 в *Handbook of Econometrics* (под редакцией Heckman, J.J. & E.E. Leamer), том 5. Elsevier Science.

Inoue, A. & L. Kilian (2002). Bootstrapping autoregressive processes with possible unit roots. *Econometrica* 70, 377–391.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of Economics & Statistics* 80, 218–230.

Kilian, L. (1999). Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses. *Review of Economics & Statistics* 81, 652–660.

Kim J.H. (1999). Asymptotic and bootstrap prediction regions for vector autoregression. *International Journal of Forecasting* 15, 393–403.

Park, J.Y. (2003). Bootstrap unit root tests. *Econometrica* 71, 1845–1895.

Politis, D.N., J.P. Romano & M. Wolf (1999). *Subsampling.* Springer.

Shao, J. & D. Tu (1995). *The Jackknife and Bootstrap.* Springer.

Sullivan, R., A. Timmermann & H. White (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance* 54, 1647–1691.

Wang, C.Y., S. Wang & R.J. Carroll (1997). Estimation in choice-based sampling with measurement error and bootstrap analysis. *Journal of Econometrics* 77, 65–86.

White, H. (2000). A reality check for data snooping. *Econometrica* 68, 1097–1126.

The basics of bootstrapping

Stanislav Anatolyev

New Economic School, Moscow, Russia

This essay is an introduction to principles and methodology of the bootstrap. The basics of bootstrap inference, resampling and asymptotic refinement are given. The narration is accompanied with clarifying examples. There is also a brief description of other methodological essays of the current issue of *Quantile* and references to non-included material.