

Бутстрап-схемы для временных рядов^{*}

Питер Бюльман[†]

ETH Zürich, Цюрих, Швейцария

В настоящем эссе представлен обзор и сравнение схем блочного, решетчатого и локального бутстрапа для временных рядов, а также рассмотрены их теоретические свойства и поведение в конечных выборках. Материал подобран с целью дать новую и объективную картину некоторых аспектов бутстрапирования временных рядов. Универсальность блочного бутстрапа противопоставляется решетчатому бутстрапу. Обсуждаются преимущества и недостатки реализации бутстрап-схем на практике, и утверждается, что решетчатый бутстрап часто превосходит блочный метод. Локальный бутстрап, предназначенный для непараметрического сглаживания, легок в реализации, но иногда работает плохо.

1 Введение

Бутстрапирование можно рассматривать как симуляции статистики или статистической процедуры на основе оцененного распределения \hat{P}_n наблюдаемых данных X_1, \dots, X_n . В случае зависимых наблюдений объект \hat{P}_n является более сложным и значительно менее очевидным, чем в основополагающей статье Эфрона (Efron, 1979) для случая независимых данных. Здесь в основном рассматриваются методы блочного, решетчатого и локального бутстрапирования, которые в определенном смысле являются непараметрическими и не привязанными к какой-либо модели. Целью является получение четкой картины о сильных и слабых сторонах различных способов бутстрапирования временных рядов. Для этого исследуются как их теоретические свойства, так и их поведение в конечных выборках. До сих пор в литературе очень мало внимания уделялось общей перспективе при сравнении различных бутстрап-схем. В этом отношении настоящий *избирательный* подход дает ценное и новое понимание вопроса и отличает наш сравнительный анализ от работ Léger, Politis & Romano (1992), Efron & Tibshirani (1993, гл. 8.5–8.6), Shao & Tu (1995, гл. 9), Li & Maddala (1996) или Davison & Hinkley (1997, гл. 8).

Получение информации из данных формализуется с помощью оценки $\hat{\theta}$, скалярной, векторной или в пространстве кривых. Для инференции необходимо оценить выборочное распределение оценки $\hat{\theta}$ или ее пивотизированной или студентизированной версии. Для временных рядов данная задача гораздо сложнее, чем в случае независимых наблюдений, и методы, основанные на аналитических выкладках, быстро сталкиваются с затруднениями. Рассмотрим, например, оценку $\hat{\theta}$, асимптотически нормально распределенную вокруг конечномерного параметра θ : при определенных условиях и предположении о стационарности ряда данных X_1, \dots, X_n ,

$$\sqrt{n}(\hat{\theta} - \theta) \Rightarrow \mathcal{N}(0, \sigma_\infty^2) \text{ при } n \rightarrow \infty. \quad (1)$$

В отличие от случая независимых наблюдений асимптотическая дисперсия σ_∞^2 является бесконечномерным объектом, представляющим собой бесконечную сумму ковариаций, которая,

^{*}Перевод Б. Гершмана и С. Анатольева. Эссе является сокращенной и измененной версией статьи Bühlmann (2002). Цитировать как: Бюльман, Питер (2007) «Бутстрап-схемы для временных рядов», Квантиль, №3, стр. 37–56. Citation: Bühlmann, Peter (2007) “Bootstrap schemes for time series,” Quantile, No.3, pp. 37–56.

[†]Адрес: Seminar für Statistik, ETH Zürich, CH-8092 Zürich, Switzerland. Электронная почта: buhlmann@stat.math.ethz.ch

вообще говоря, не поддается оцениванию со скоростью сходимости $1/\sqrt{n}$. В простом случае, когда $\hat{\theta} = \bar{X}_n$,

$$\sigma_\infty^2 = \sum_{k=-\infty}^{\infty} \mathbb{C}[X_0, X_k].$$

Асимптотической дисперсией в этом примере является (нормированное на 2π) значение функции спектральной плотности процесса, порождающего данные, в нуле. В другом примере, когда $\hat{\theta} = \text{med}(X_1, \dots, X_n)$,

$$\sigma_\infty^2 = \sum_{k=-\infty}^{\infty} \mathbb{C}[\text{IF}(X_0), \text{IF}(X_k)], \quad \text{IF}(x) = \frac{\text{sign}(x - \theta)}{2f(\theta)},$$

где $\theta = F^{-1}(1/2)$ – медиана маржинального распределения F для X_t , имеющего функцию плотности f . Здесь фигурирует спектральная плотность процесса с функциями влияния $(\text{IF}(X_t))_{t \in \mathbb{Z}}$, то есть мгновенное неизвестное преобразование процесса $(X_t)_{t \in \mathbb{Z}}$. Было бы очень неудобно оценивать неизвестную функцию плотности f и θ для получения оценки $\text{IF}(\cdot)$, а затем ее спектральной плотности. Преимущество бутстрапирования в том, что оно позволяет состоятельно оценивать асимптотическую дисперсию и распределение $\sqrt{n}(\hat{\theta} - \theta)$ *автоматически*.

Состоятельность, или «точность первого порядка», заключается в требовании состоятельного оценивания предельного распределения $\hat{\theta}$. Точнее говоря, для оценки $\hat{\theta}$ из пространства \mathbb{R}^q

$$\sup_{x \in \mathbb{R}^q} |\mathbb{P}^*[a_n(\hat{\theta}^* - \theta^*) \leq x] - \mathbb{P}[a_n(\hat{\theta} - \theta) \leq x]| = o_P(1) \quad (n \rightarrow \infty), \quad (2)$$

где $a_n(\hat{\theta} - \theta)$ сходится к невырожденному предельному распределению. Символ ‘ \leq ’ определен покомпонентно, и, как обычно, звездочка * обозначает бутстраповский аналог. Центральное значение θ^* , являющееся константой условно на исходных наблюдениях X_1, \dots, X_n , обычно не выбирается равным $\hat{\theta}$, как в случае Эфроновского бутстрапа для случая независимых одинаково распределенных данных; подробности даны ниже при обсуждении конкретных методов бутстрапирования временных рядов.

Например, если $\hat{\theta}$ является выборочным средним или медианой, предельное распределение имеет вид $\mathcal{N}(0, \sigma_\infty^2)$, как в (1), с дисперсией σ_∞^2 , приведенной выше (при выполнении некоторых условий регулярности). Состоятельность тогда следует из того, что

$$\sqrt{n}(\hat{\theta}^* - \theta^*) \Rightarrow \mathcal{N}(0, \sigma_\infty^2) \text{ по вероятности при } n \rightarrow \infty,$$

то есть предельные распределения бутстрапованной и исходной оценок совпадают. Эта сходимость обычно требует среди прочего, чтобы бутстраповская дисперсия была асимптотически верной, то есть

$$n\mathbb{V}^*(\hat{\theta}^*) = \sigma_\infty^2 + o_P(1) \quad (n \rightarrow \infty).$$

Поскольку $\sigma_\infty^2 = \lim_n n\mathbb{V}(\hat{\theta})$, данное требование можно рассматривать как сходимость нормированных дисперсий

$$n\mathbb{V}^*(\hat{\theta}^*) - n\mathbb{V}(\hat{\theta}) = o_P(1) \quad (n \rightarrow \infty).$$

Состоятельность бутстрапа (2) обычно имеет место, если $\hat{\theta}$ асимптотически нормальна. Помимо аппроксимации распределения оценки $\hat{\theta}$, бутстрап позволяет аппроксимировать $\mathbb{V}(\hat{\theta})$ бутстраповской дисперсией $\mathbb{V}^*(\hat{\theta}^*)$. Точность оценивания распределения в (2) зависит от точности бутстраповской дисперсии

$$a_n^2 \mathbb{V}^*(\hat{\theta}^*) - a_n^2 \mathbb{V}(\hat{\theta}), \quad (3)$$

при условии, что существуют разложения Эджворта для $\hat{\theta}$ и $\hat{\theta}^*$. Обычно бесконечномерность предельной дисперсии делает задачу точного оценивания бутстраповской дисперсии гораздо сложнее, чем в случае независимых наблюдений.

Конечно, как и в случае независимых данных, бутстрапирование временных рядов дает преимущество в виде точности приближения более высокого порядка по сравнению с аппроксимацией с помощью нормального распределения как в (1). Для этого приближение оценивается для студентизированных версий $\hat{\theta}$, или же доверительный интервал корректируется с помощью BC_a [коррекция на смещение и увеличение скорости сходимости (Efron, 1987)] или двойного бутстрапа. Но в случае *конечной* выборки схемы первого порядка часто могут быть такими же точными, как их аналоги второго порядка, и (3) желательно надлежащим образом ограничить. Значительная часть настоящей статьи посвящена обсуждению точности первого порядка, но также затрагиваются аспекты приближения второго порядка.

2 Блочный бутстрап

Блочный бутстрап пытается имитировать поведение оценки $\hat{\theta}$ с помощью случайного независимого ресэмплинга $X_{t+1}, \dots, X_{t+\ell}$ последовательных наблюдений: группировка в блоки используется для сохранения исходной структуры временного ряда внутри блока. Подобная идея появилась в статье Hall (1985), но прорывом в блочном бутстрапе стала статья Künsch (1989), где в деталях объясняется, как и почему такая бутстрап-схема работает.

2.1 Процедура блочного бутстрапа

Правильное применение схемы блочного бутстрапа включает, во-первых, адаптацию к задаче. Предположим, что статистика $\hat{\theta}$ оценивает параметр θ , являющийся функционалом m -мерного маргинального распределения временного ряда. Например, автокорреляция первого порядка $\text{Cov}(X_0, X_1)$ для стационарного временного ряда является функционалом распределения (X_0, X_1) , что соответствует $m = 2$. Рассмотрим далее вектора последовательных наблюдений

$$Y_t = (X_{t-m+1}, \dots, X_t), \quad t = m, \dots, n \quad (4)$$

и создадим блочно-бутстраповскую выборку на основе этих векторизованных наблюдений следующим образом. Построим пересекающиеся блоки последовательных векторов $(Y_m, \dots, Y_{m+\ell-1}), (Y_{m+1}, \dots, Y_{m+\ell}), \dots, (Y_{n-\ell+1}, \dots, Y_n)$, где $\ell \in \mathbb{N}$ – длина блока. Для простоты сначала предположим, что число блоков $n - m + 1 = k\ell$, $k \in \mathbb{N}$. Далее осуществляем независимые вытягивания k блоков с возвращением,

$$Y_{S_1+1}, \dots, Y_{S_1+\ell}, Y_{S_2+1}, \dots, Y_{S_2+\ell}, \dots, Y_{S_k+1}, \dots, Y_{S_k+\ell}, \quad (5)$$

где точки начала блоков S_1, \dots, S_k независимо равномерно распределены на множестве $\{m - 1, \dots, n - \ell\}$ всех возможных начальных точек. Если число блоков $n - m + 1$ не кратно ℓ , ресэмплируют $k = \lfloor (n - m + 1) / \ell \rfloor + 1$ блоков, но от k -го блока используется только часть, чтобы в итоге получить $n - m + 1$ m -мерных векторов. Таким образом сформированную последовательность блоков m -мерных векторов в (5) можно назвать блочно-бутстраповской выборкой. Однако, как будет видно далее, блочно-бутстрапованная оценка не определяется принципом подстановки, и понятие бутстраповской выборки тогда не так прозрачно.

Хорошо определить блочно-бутстрапованную оценку не совсем просто. Векторизация в (4) обычно так соотносится с оценкой, что $\hat{\theta}$ симметрична относительно векторизованных наблюдений Y_m, \dots, Y_n . Например, часто оценку можно представить в виде

$$\hat{\theta} = T(F_n^{(m)}), \quad (6)$$

где $F_n^{(m)}(\cdot) = (n - m + 1)^{-1} \sum_{t=m}^n \mathbb{I}_{[Y_t \leq \cdot]}$ — эмпирическая функция распределения m -мерного маржинального распределения $(X_t)_{t \in \mathbb{Z}}$, а T — гладкий функционал.

Пример А. Для автокорреляции первого порядка $\theta = \text{Cov}(X_0, X_1)$ рассмотрим оценку $\hat{\theta} = \hat{R}(1)/\hat{R}(0)$, где $\hat{R}(k) = (n - 1)^{-1} \sum_{t=1}^{n-1} (X_t - \hat{\mu}_X)(X_{t+k} - \hat{\mu}_X)$ ($k \in \{0, 1\}$), $\hat{\mu}_X = (n - 1)^{-1} \sum_{t=1}^{n-1} X_t$. Эта оценка $\hat{\theta}$ симметрична относительно Y_2, \dots, Y_n , где $Y_t = (X_{t-1}, X_t)$, и имеет вид (6), причем $m = 2$. Отметим, что обычная оценка — это $\theta = \tilde{R}(0)/\tilde{R}(1)$, где $\tilde{R}(k) = n^{-1} \sum_{t=1}^n (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n)$, что приблизительно равняется $\hat{\theta}$, если забыть об эффектах на краях.

Пример В. GM-оценки (обобщенные M-оценки) в AR(p)-модели можно записать в виде (6) для $m = p + 1$. Они определяются неявно, по аналогии с нормальными уравнениями:

$$\sum_{t=p+1}^n w_t \psi((X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_p X_{t-p}) \sigma^{-1})(X_{t-1}, \dots, X_{t-p})^T = 0,$$

где $\psi : \mathbb{R} \rightarrow \mathbb{R}$, σ^2 — дисперсия инноваций, а $(w_t)_{t=p+1}^n$ — последовательность соответствующих весов; см. Martin & Yohai (1986). Помимо гауссовских ММП-оценок с $\psi(x) = x$, они включают оценки, робастные к выбросам в инновациях и лагированных значениях.

Блочно-бутстрапированная оценка, соответствующая (6), определяется как

$$\hat{\theta}^{*B} = T(F_n^{(m)*B}), \quad F_n^{(m)*B}(\cdot) = (n - m + 1)^{-1} \sum_{i=1}^k \sum_{t=S_i+1}^{S_i+\ell} \mathbb{I}_{[Y_t \leq \cdot]}, \quad (7)$$

где k и S_i означают то же, что и в (5). Центральное значение θ^* для блочного бутстрапа в (2) часто равняется $\mathbb{E}^{*B}[\hat{\theta}^{*B}]$, что, вообще говоря, отлично от $\hat{\theta}$. Это определение блочно-бутстрапированной оценки, данное в Künsch (1989), можно интерпретировать следующим образом. Если $\hat{\theta} = g_{n-m+1}(Y_m, \dots, Y_n)$ является симметричной функцией $g_{n-m+1}(\cdot)$ от $n - m + 1$ векторизованных наблюдений, то

$$\hat{\theta}^{*B} = g_{n-m+1}(Y_{S_1+1}, \dots, Y_{S_1+\ell}, Y_{S_2+1}, \dots, Y_{S_2+\ell}, \dots, Y_{S_k+1}, \dots, Y_{S_k+\ell}),$$

где использован принцип подстановки для векторизованных наблюдений. В частности, блочно-бутстрапированная оценка определяется для таких Y , которые попадают в множество исходных векторизованных наблюдений. Это будет *не так* без векторизации (4). Рис. 1 иллюстрирует недостаток наивного блочного бутстрапа, использующего $m = 1$ вместо правильного $m = 2$ в примере А. Наиболее яркий дефект псевдовыборки, полученной с помощью наивного блочного бутстрапа, состоит в появлении новых точек внутри прямоугольников в верхнем левом и нижнем правом углах диаграммы. Наивная блочно-бутстрапированная оценка (например, для автокорреляции в примере А), которая использует принцип подстановки с выборкой, полученной с помощью наивного блочного бутстрапа, может быть сильно подпорчена этими вновь созданными плохими точками. Как уже было сказано, этот недостаток отсутствует в определении блочного бутстрапа на основе векторизованных наблюдений (7).

Касательно процедуры блочного бутстрапа остаются как минимум две трудности, которые необходимо рассматривать в каждом отдельном случае.

- (1) Из-за неприменимости принципа подстановки часто необходимо изменение способа *вычисления* $\hat{\theta}^{*B}$, что может быть очень неудобно.
- (2) Векторизация как в (4) не всегда адекватна. Например, MA-параметр в MA(1)-модели или спектральная плотность стационарного процесса зависят от всего распределения процесса, что соответствует $m = \infty$.

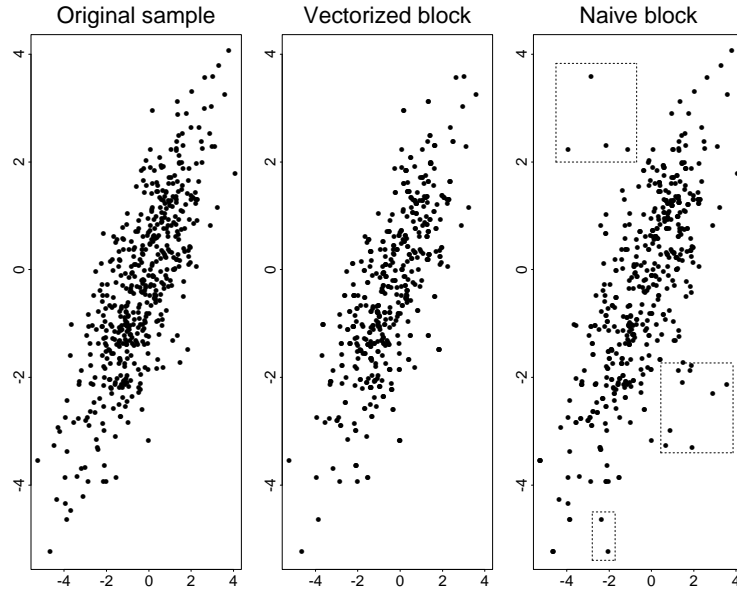


Рис. 1: Диаграммы пар наблюдений, отстоящих на один шаг, в бутстраповских выборках размера $n = 512$. Слева: исходная выборка (X_{t-1}, X_t) , $t = 2, \dots, n$. В центре: псевдовыборка для блочного бутстрапа Y_{S_i+j} , $i = 1, \dots, k = 64$, $j = 1, \dots, \ell = 8$ из (5) для $m = 2$. Справа: псевдовыборка для наивного блочного бутстрапа $(X_{t-1}^{*nB}, X_t^{*nB})$, $t = 2, \dots, n$, где X_t^{*nB} – последовательное t -ое значение из (5) для $m = 1$, $k = 64$, $\ell = 8$; точки внутри прямоугольников (и не только) не встречаются на диаграмме слева.

В тех случаях, когда проблема (1) или (2) становится слишком затруднительной, решение по умолчанию – игнорировать шаг (4), на котором происходит векторизация, и работать с наивным блочным бутстрапом (используя $m = 1$). В результате может произойти значительное снижение эффективности метода. Предложения для решения проблемы (2), главным образом для случая оценивания спектральной плотности, даны в Politis & Romano (1992) и Bühlmann & Künsch (1995).

2.2 Область применимости и точность

Блочный бутстрап предназначен для работы с общими стационарными процессами, порождающими данные, $(X_t)_{t \in \mathbb{Z}}$, где $X_t \in \mathbb{R}^d$ ($d \geq 1$) или принимает значения в пространстве категорий. С точки зрения асимптотики, длина блока ℓ должна расти при $n \rightarrow \infty$, но не слишком быстро. Если ограничиваться только процессами с коротким горизонтом зависимости (например, суммируемые автоковариации или коэффициенты перемешивания), применение блочного бутстрапа теоретически оправдано во многих обстоятельствах, например, для таких оценок, как в (6) с гладким T , см. Künsch (1989), Bühlmann (1994). Другие ссылки даны в разделе 7. Для случаев долгосрочной зависимости разработана некоторая теория и модификации для случая $\hat{\theta} = \bar{X}_n$: Lahiri (1993) показывает, что блочный бутстрап состоятелен, если \bar{X}_n имеет нормальное предельное распределение, но бутстрапированная статистика должна быть скорректирована на коэффициент, зависящий от обычно неизвестной скорости сходимости, например от параметра автомоделирования в автомоделных процессах. Если \bar{X}_n не имеет нормального распределения в качестве предельного из-за долгосрочной зависимости, Hall, Jing & Lahiri (1998) показывают состоятельность модифицированного метода блочных подвыборок. Когда наблюдения имеют маргинальное распределение с тяжелыми хвостами, в Lahiri (1995) показано, что блочный бутстрап с размером псевдовыборки $m \ll n$ состоятелен для случая $\hat{\theta} = \bar{X}_n$.

Что касается точности блочного бутстрапа, рассмотрим вначале оценивание асимптотической дисперсии $\hat{\theta}$. Künsch (1989) доказал, что для среднего квадрата ошибки (MSE)

$$\mathbb{E}[(n\mathbb{V}^{*B}(\hat{\theta}^{*B}) - n\mathbb{V}(\hat{\theta}))^2] \propto n^{-2/3}, \quad (8)$$

что достигается при оптимальной с точки зрения скорости сходимости длине блока $\ell \propto n^{1/3}$. Заметим, что это соответствует (3) при $a_n = \sqrt{n}$. Основные предположения для этого результата требуют, чтобы $n\mathbb{V}(\hat{\theta})$ сходилась к невырожденной предельной дисперсии, T в (6) был достаточно гладким, а также выполнения некоторых условий перемешивания для стационарного процесса $(X_t)_{t \in \mathbb{Z}}$. Несколько удивительно, что скорость $n^{-2/3}$ не зависит от «степени зависимости», то есть того, как быстро автокорреляции, или, более общо, коэффициенты перемешивания, падают по мере увеличения лага между наблюдениями. В частности, даже если автоковариации и коэффициенты перемешивания убывают экспоненциально, скорость для MSE все равно равна $n^{-2/3}$. Таким образом, оценка дисперсии с помощью блочного бутстрапа не адаптирует скорость сходимости к степени зависимости соответствующего процесса. Объяснение отсутствию этой адаптивности было дано уже в Künsch (1989): оценка дисперсии блочным бутстрапом асимптотически эквивалентна оценке спектральной плотности в нуле с треугольным окном:

$$n\mathbb{V}^{*B}(\hat{\theta}^{*B}) \approx \sum_{k=-\ell}^{\ell} \left(1 - \frac{|k|}{\ell}\right) \hat{R}_{\text{IF}}(k), \quad (9)$$

где $\hat{R}_{\text{IF}}(k)$ – эмпирическая ковариация $(\text{IF}(Y_t; F^{(m)}))_{t=m}^n$ порядка k , где $\text{IF}(\cdot; F^{(m)})$ – функция влияния оценки при соответствующем истинном m -мерном маргинальном распределении $F^{(m)}$; функция влияния $\text{IF}(\cdot; F^{(m)})$ – это преобразование, которое асимптотически линейризует подходяще регулярную оценку, см. (10). Но треугольная форма окна $1 - |k|/\ell$ ($k = -\ell, \dots, 0, \dots, \ell$) делает невозможным улучшение по сравнению со скоростью MSE $n^{-2/3}$. Конический блочный бутстрап, предложенный в Künsch (1989, формула (2.12)) и более подробно изученный в Paparoditis & Politis (2001), преодолевает это ограничение.

Что касается построения доверительных областей, Götze & Künsch (1996) доказали, что распределение надлежащим образом определенной студентизированной версии $\hat{\theta}$ можно аппроксимировать блочным бутстрапом с точностью, близкой к $O_P(n^{-2/3})$, используя длину блока $\ell \propto n^{1/3}$.¹ Как и при оценивании дисперсии, степень точности не может быть улучшена для временных рядов с геометрически убывающими свойствами зависимости. Götze & Künsch (1996) также представляют модификацию предложенной Эфроном (Efron, 1987) корректировки BC_a (удаление смещения и увеличение скорости сходимости). В конечных выборках метод, точный во втором приближении, не всегда может быть выгодным. К сожалению, нелегко судить по данным о целесообразности применения техники второго порядка. Двойное блочное бутстрапирование для корректировки доверительной области, полученной с помощью бутстрапа первого порядка, непросто, так как зависимость нарушается в местах соединения блоков, см. Davison & Hall (1993) и Choi & Hall (2000).

2.3 Выбор длины блока ℓ

Оптимальная длина блока, будучи параметром настройки блочного бутстрапа, зависит по крайней мере от трех вещей: процесса, порождающего данные, бутстрапируемой статистики и цели использования бутстрапа, например, оценивания смещения, дисперсии или распределения.

¹Эту скорость можно улучшить до приблизительно $O_P(n^{-3/4})$ при использовании для студентизации оценки дисперсии, которая принимает отрицательные значения с положительной вероятностью.

Рассмотрим сначала оценивание с помощью блочного бутстрапа дисперсии оценки $\hat{\theta}$ вида (6). В этом случае

$$\hat{\theta} \approx (n - m + 1)^{-1} \sum_{t=m}^n \text{IF}(Y_t; F^{(m)}), \quad (10)$$

где $\text{IF}(\cdot; F^{(m)})$ – функция влияния $\hat{\theta}$ в $F^{(m)}$. На основе этой линеаризации формулу (9) можно переписать в виде

$$n \mathbb{V}^{*B}(\hat{\theta}^{*B}) \approx 2\pi \hat{f}_{\text{IF}}(0), \quad (11)$$

где $\hat{f}_{\text{IF}}(\lambda)$ ($0 \leq \lambda \leq \pi$) – оценка спектральной плотности с треугольным окном при частоте λ и ширине окна ℓ^{-1} , основанная на функциях влияния $(\text{IF}(Y_t; F^{(m)}))_{t=m}^n$. Таким образом, длина блока имеет интересную интерпретацию как величина, обратная к ширине окна при оценке спектральной плотности. Отсюда следует, что асимптотически оптимальная в смысле MSE длина блока для оценки дисперсии есть

$$\ell_{opt} \propto n^{1/3}.$$

Bühlmann & Künsch (1999) предлагают оценивать ℓ_{opt} с помощью итеративной схемы подстановки для выбора локально оптимальной ширины окна при оценке спектральной плотности при частоте 0, используя асимптотическую эквивалентность в (11).

Более общий метод, который к тому же применим для выбора оптимальной длины блока ℓ при оценивании распределения, был предложен в Hall, Horowitz & Jing (1995). Авторы рассматривают поведение блочного бутстрапа при разных значениях длины блоков для подвыборок размера $m \ll n$ и получают оптимальную длину блока для размера подвыборки m . Затем оцененная оптимальная длина блока выводится путем экстраполяции Ричардсона до размера исходной выборки n . Этот метод требует спецификации размера подвыборки m , что менее критично, чем выбор длины блока. Подобная техника использования подвыборок является очень общей, но может быть не очень эффективной. В частности, если оценка $\hat{\theta}$ сильно нелинейная, свойства метода в подвыборках могут быть очень плохими; подобная неэффективность в аналогичных обстоятельствах иллюстрируется в подразделе 4.2.

Что касается оценки смещения с помощью блочного бутстрапа, Lahiri (1999) показывает, что асимптотически оптимальные в смысле MSE длины блоков одинаковы для оценивания смещения и дисперсии: оцененные длины блоков для дисперсии могут, таким образом, быть использованы при оценивании смещения.

Автоматический выбор длины блока по крайней мере так же труден в контексте временных рядов, как выбор *локального* параметра настройки наподобие ширины окна. Хуже того, формула (11) указывает на эквивалентность задаче выбора ширины окна только в асимптотике: линеаризация в (10) может иметь значительный эффект в конечных выборках. Более того, длина блока ℓ не имеет практической интерпретации, а диагностические инструменты для ее выбора пока что не разработаны.

3 AR-решетчатый бутстрап для стационарных линейных временных рядов

Решетчатый бутстрап основан на идее решетчатого приближения (Grenander, 1981) процесса $(X_t)_{t \in \mathbb{Z}}$ классом (полу)параметрических моделей. В таком случае бутстрап является не чем иным, как симуляцией из оцененного с помощью решетки процесса.

Временной ряд называется линейным и обратимым, если он представим в виде авторегрессии бесконечного порядка $\text{AR}(\infty)$,

$$X_t - \mu_X = \sum_{j=1}^{\infty} \phi_j (X_{t-j} - \mu_X) + \epsilon_t \quad (t \in \mathbb{Z}), \quad (12)$$

где $\mu_X = \mathbb{E}[X_t]$, $(\epsilon_t)_{t \in \mathbb{Z}}$ – последовательность IID-инноваций, $\mathbb{E}[\epsilon_t] = 0$, и ϵ_t не зависят от $\{X_s; s < t\}$. Это определение корректно, если, например, $\mathbb{E}[\epsilon_t^2] < \infty$ и $\sum_{j=1}^{\infty} \phi_j^2 < \infty$.

3.1 Процедура AR-решетчатого бутстрапа

AR-решетчатое приближение строится с помощью AR(p) моделей

$$X_t - \mu_X = \sum_{j=1}^p \phi_j (X_{t-j} - \mu_X) + \epsilon_t \quad (t \in \mathbb{Z}),$$

где μ_X и ϵ_t определены как в (12). Для исходных данных сначала выбирается порядок авторегрессии \hat{p} , например, с помощью критерия AIC для гауссовских инноваций, см. Shibata (1980). Оставшийся интересующий исследователя объект $\eta_{\hat{p}} = (\mu_X, (\phi_1, \dots, \phi_{\hat{p}}), F_{\epsilon})$ является полупараметрическим. Здесь F_{ϵ} обозначает распределение IID-инноваций ϵ_t . Оценки выбираются следующим образом:

$$\hat{\mu}_X = 1/n \sum_{t=1}^n X_t, \quad (\hat{\phi}_1, \dots, \hat{\phi}_{\hat{p}}) \text{ по методу Юла-Уолкера,}$$

$$\hat{F}_{\epsilon}(x) = \hat{\mathbb{P}}[\epsilon_t \leq x] = (n - \hat{p})^{-1} \sum_{t=\hat{p}+1}^n \mathbb{I}_{[R_t - \bar{R}_{\bullet} \leq x]}, \quad R_t = X_t - \sum_{j=1}^{\hat{p}} \hat{\phi}_j X_{t-j},$$

где \bar{R}_{\bullet} – среднее доступных остатков R_t ($t = \hat{p} + 1, \dots, n$).

Оценки $\hat{p}, \hat{\eta}_{\hat{p}}$ характеризуют распределение $\hat{P}_{n; \text{AR}}$ для авторегрессионного процесса. Оно может быть представлено в виде следующего AR(\hat{p})-уравнения:

$$X_t^{*\text{AR-S}} - \hat{\mu}_X = \sum_{j=1}^{\hat{p}} \hat{\phi}_j (X_{t-j}^{*\text{AR-S}} - \hat{\mu}_X) + \epsilon_t^* \quad (t \in \mathbb{Z}), \quad (13)$$

где $(\epsilon_t^*)_{t \in \mathbb{Z}}$ – последовательность IID-инноваций, с маржинальным распределением $\epsilon_t^* \sim \hat{F}_{\epsilon}$.

Тогда AR-решетчатая бутстраповская выборка – это конечная выборка X_1^*, \dots, X_n^* из процесса (13), имеющего распределение $\hat{P}_{n; \text{AR}}$. На практике вычисление происходит так. Начиная с $(X_{-u}^*, \dots, X_{-u+\hat{p}-1}^*) = (\hat{\mu}_X, \dots, \hat{\mu}_X)$ при большом u , например, $u = 1000$. Далее симулируем X_t^* для $t = -u + \hat{p}, \dots, 0, 1, \dots, n$ в соответствии с (13). Поскольку процесс в (13) является (с большой вероятностью) марковским и с геометрическим перемешиванием, значения X_1^*, \dots, X_n^* из симулированной выборки являются очень хорошей аппроксимацией для выборки из стационарного распределения процесса (13). Далее AR-решетчато-бутстрапированная оценка $\hat{\theta}^{*\text{AR-S}}$ строится с помощью правила подстановки. Записав $\hat{\theta} = h_n(X_1, \dots, X_n)$ как функцию от исходных данных X_1, \dots, X_n , определим

$$\hat{\theta}^{*\text{AR-S}} = h_n(X_1^{*\text{AR-S}}, \dots, X_n^{*\text{AR-S}}). \quad (14)$$

Подобная схема бутстрапа была предложена в Kreiss (1992) и впоследствии исследована в Bühlmann (1997), Bickel & Bühlmann (1999) и Choi & Hall (2000).

Центральное значение θ^* в (2) для AR-решетчатого бутстрапа находится следующим образом. Параметр θ является функционалом истинного процесса $(X_t)_{t \in \mathbb{Z}} \sim P$, тогда $\theta^{*\text{AR-S}}$ – тот же самый функционал, вычисленный для оцененного распределения $\hat{P}_{n; \text{AR}}$, которое генерирует бутстрапированный процесс в (13).

Пример А (продолжение). Для оценки автокорреляции первого порядка $\theta^{*\text{AR-S}} = \text{Corr}^{*\text{AR-S}}(X_t^{*\text{AR-S}}, X_{t+1}^{*\text{AR-S}})$.

Заметим, что, вообще говоря, $\mathbb{E}^{*\text{AR-S}}[\hat{\theta}^{*\text{AR-S}}] \neq \theta^{*\text{AR-S}}$. Вычисление $\theta^{*\text{AR-S}}$ можно осуществить с помощью быстрой оценки Монте-Карло:

- (1) Сгенерировать одну очень длинную реализацию $X_1^{*AR-S}, \dots, X_v^{*AR-S}$, $v \gg n$.
- (2) Использовать $\hat{\theta}_v^{*AR-S} = h_v(X_1^{*AR-S}, \dots, X_v^{*AR-S})$ в качестве Монте-Карло аппроксимации для θ^{*AR-S} .

Оправдание аппроксимации на шаге (2) дает формула (2), означающая, что $\hat{\theta}_v^{*AR-S}$ сходится к θ^{*AR-S} со скоростью $a_v^{-1} \ll a_n^{-1}$ (в предположении, что $a_n(\hat{\theta}_n - \theta)$ сходится к невырожденному распределению).

3.2 Область применимости и точность

Процедура AR-решетчатого бутстрапа сильно полагается на предположение о том, что данные X_1, \dots, X_n являются конечной реализацией процесса $AR(\infty)$ из (12). В такой постановке состоятельность $\hat{\theta}$ в (2), являющейся гладкой функцией от средних, доказана в Bühlmann (1997); этот результат расширен в Bickel & Bühlmann (1999) для $\hat{\theta}$, имеющей вид (6). Согласно этим результатам, аппроксимирующий порядок авторегрессии должен асимптотически расти, хотя не слишком быстро, при $n \rightarrow \infty$. Представление $AR(\infty)$ включает важный класс ARMA-моделей

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + \sum_{k=1}^q \psi_k \epsilon_{t-k} + \epsilon_t \quad (t \in \mathbb{Z})$$

с обратимым генерирующим MA-полиномом, т.е. корни $\Psi(z) = 1 + \sum_{k=1}^q \psi_k z^k$ ($z \in \mathbb{C}$) лежат вне единичного круга $\{z \in \mathbb{C}; |z| \leq 1\}$. Здесь $(\epsilon_t)_{t \in \mathbb{Z}}$ – последовательность IID-инноваций; также должны выполняться некоторые дополнительные условия регулярности, стандартные для теории ARMA-моделей. Но, конечно, существует множество процессов, не представимых в виде $AR(\infty)$, например, приведенные ниже нелинейная авторегрессия второго порядка $AR(2)$, см. (19), или билинейная модель (20). К сожалению, тестирование на линейность или представимость в виде $AR(\infty)$ весьма деликатно: Bickel & Bühlmann (1997) показывают, что замыкание линейных или AR-процессов на удивление велико. Этот факт отражает сложность суждения по конкретному набору данных о том, годится ли для применения AR-решетчатый бутстрап.

В классе линейных обратимых временных рядов, определенных в (12), AR-решетчатый бутстрап имеет высокую точность: теоретические и практические исследования показывают, что он обычно превосходит более общую процедуру блочного бутстрапа из раздела 2. В Bühlmann (1997) показано, что для $\hat{\theta} = \bar{X}_n$ и при аппроксимации порядка авторегрессии \hat{p} с помощью критерия AIC,

$$n\mathbb{V}^{*AR-S}(\bar{X}_n^{*AR-S}) - n\mathbb{V}(\bar{X}_n) = O_P(n^{-(v-2)/(2v)}),$$

если истинные параметры авторегрессии $(\phi_j)_{j \in \mathbb{N}}$ убывают как $\phi_j \leq const \cdot j^{-v}$ ($v > 2$). В частности, если ϕ_j убывают экспоненциально, то

$$n\mathbb{V}^{*AR-S}(\bar{X}_n^{*AR-S}) - n\mathbb{V}(\bar{X}_n) = O_P(n^{-1/2+\kappa}) \text{ для любого } \kappa > 0. \quad (15)$$

Эти два результата показывают, что метод *автоматически адаптируется* к характеру убывания соответствующей структуры зависимости, что является весьма желательным свойством, которым не обладает блочный бутстрап, см. (8). Эти результаты об адаптивности важны не только в асимптотике, но также могут быть использованы при симуляциях в конечных выборках, см. подраздел 4.1.

AR-решетчатый бутстрап очень точен не только при оценивании дисперсии: Choi & Hall (2000) выводят свойства второго порядка для построения доверительных областей. Они предлагают калибровать полученную область первого порядка с помощью двойного бутстрапирования, основываясь на идеях, берущих начало в работах Hall (1986), Veran (1987) и Loh

(1987). Рассмотрим построение двустороннего доверительного интервала, покрывающего θ с вероятностью $1 - \alpha$. Интервал первого порядка имеет вид $[\hat{\theta} - \hat{r}_{1-\alpha/2}, \hat{\theta} - \hat{r}_{\alpha/2}]$, где \hat{r}_{α} – α -квантиль распределения $\hat{\theta}^{*AR-S} - \theta^{*AR-S}$ условно на X_1, \dots, X_n .

Теперь рассмотрим дополнительную корректировку исходного номинального уровня покрытия с использованием двойного бутстрапа. На основе $X_1^{*AR-S}, \dots, X_n^{*AR-S}$ применяем AR-решетчатый бутстрап и получаем $X_1^{**AR-S}, \dots, X_n^{**AR-S}$. Тогда \hat{r}_{α}^{*AR-S} – α -квантиль распределения $\hat{\theta}^{**AR-S} - \theta^{**AR-S}$ условно на $X_1^{*AR-S}, \dots, X_n^{*AR-S}$.

Пусть

$$\hat{a}(1 - q) = \mathbb{P}^{*AR-S}[\hat{\theta}^{*AR-S} - \hat{r}_{1-q/2}^{*AR-S} \leq \theta^{*AR-S} \leq \hat{\theta}^{*AR-S} - \hat{r}_{q/2}^{*AR-S}], \quad (16)$$

что измеряет действительную степень покрытия при номинальном уровне $1 - q$ для бутстрапа второго уровня, основанного на бутстрапированных данных первого уровня $[\theta^{*AR-S} - \text{константа, зависящая только от } X_1, \dots, X_n]$. Далее рассмотрим $\hat{s}_{1-\alpha} = \hat{a}^{-1}(1 - \alpha)$, т.е. $(1 - \alpha)$ -квантиль \hat{a} , рассматриваемой как функция распределения, который корректирует номинальный уровень покрытия $1 - \alpha$ до $\hat{s}_{1-\alpha}$. Теперь берем

$$[\hat{\theta} - \hat{r}_{\{1-(1-\hat{s}_{1-\alpha})/2\}}, \hat{\theta} - \hat{r}_{\{(1-\hat{s}_{1-\alpha})/2\}}] \quad (17)$$

в качестве двустороннего, полученного с помощью двойного бутстрапа доверительного интервала для θ с номинальным уровнем покрытия $1 - \alpha$. Как показано в Choi & Hall (2000), этот интервал второго порядка корректен. Отметим, что явное (сложное) оценивание дисперсии при зависимых данных не является необходимым для студентизации. Choi & Hall (2000) приводят результаты симуляций, которые показывают, что этот интервал второго порядка может дать значительное улучшение и «никогда» не оказывается значительно хуже построения первого порядка.

3.3 Выбор аппроксимирующего порядка авторегрессии

Предлагается использовать AR-решетчатую аппроксимацию вместе с процедурой выбора модели с помощью критерия AIC при гауссовских инновациях. Shibata (1980) доказал оптимальность критерия AIC для прогнозирования в AR(∞)-моделях. Более того, формула (15) и предшествующая ей основаны на AIC и объясняют, почему этот критерий является хорошим выбором при оценивании дисперсии $\hat{\theta} = \bar{X}_n$.

По аналогии с задачей выбора оптимальной длины блока в блочном бутстрапе, оптимальный порядок авторегрессии, вообще говоря, зависит от истинного процесса, порождающего данные, бутстрапируемой статистики и цели применения бутстрапа. Критерий AIC имеет приятное свойство автоматически выбирать более высокий порядок для моделей с большей степенью зависимости. Ничего неизвестно о том, как адаптировать выбор порядка в AR-решетчатой аппроксимации к бутстрапируемой статистике или различным случаям оценивания дисперсии или распределения с помощью бутстрапа.

Параметр настройки процедуры AR-решетчатого бутстрапа, а именно выбор AR-модели, имеет интерпретацию и позволяет проводить диагностическую проверку, включая графические процедуры для AR-остатков. Это отличает его от параметров настройки типа ширины окна, таких как длина блока в подразделе 2.3, которые не имеют интерпретации и нелегко диагностируются по данным. Наш практический опыт говорит о том, что выбор аппроксимирующего порядка авторегрессии достаточно *нечувствителен* в плане свойств AR-решетчатого бутстрапа при условии, что выбранный порядок разумен.

4 Блочный и AR-решетчатый бутстрап в действии

4.1 AR-решетчатый и блочный бутстрап для симулированных временных рядов

Для сравнения двух схем бутстрапа мы рассматриваем симуляционные эксперименты с двумя разными процессами, но для одной и той же статистики – выборочной медианы $\hat{\theta} = \text{med}(X_1, \dots, X_n)$, представителя простых нелинейных оценок. Размер выборки $n = 512$. Кроме того, параметр настройки для AR-решетчатого бутстрапа выбирается путем минимизации критерия AIC; $\ell = 8 = n^{1/3}$ в соответствии с простым правилом выбора оптимальной асимптотической скорости для оценивания дисперсии, а $\hat{\ell}$ – из Bühlmann & Künsch (1999) для блочно-бутстраповской дисперсии, как указано в подразделе 2.3.

В первом эксперименте рассмотрим линейный ARMA(1, 1)-процесс

$$X_t = -0.8X_{t-1} - 0.5\epsilon_{t-1} + \epsilon_t, \quad (18)$$

где $(\epsilon_t)_{t \in \mathbb{Z}}$ – IID-последовательность, независимая от $\{X_s; s < t\}$, $\epsilon_t \sim t_6$. Эта модель представима в виде AR(∞)-процесса (12).

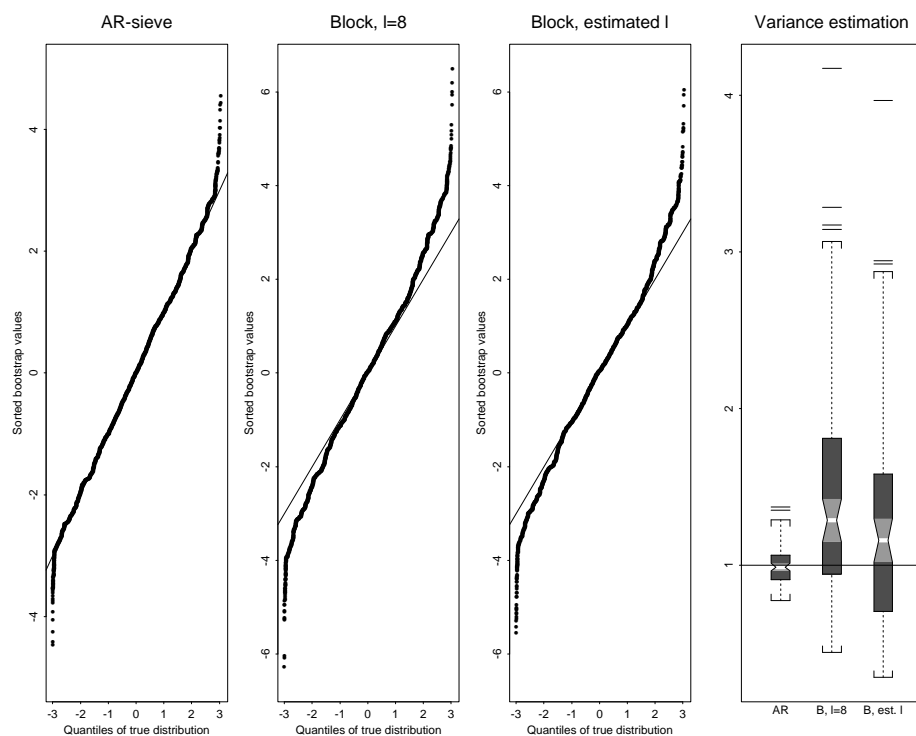


Рис. 2: Линейная модель (18), $n = 512$. Бутстраповское распределение и оценивание дисперсии $(\hat{\theta} - \mathbb{E}[\hat{\theta}])/\sigma_n$ с помощью $(\hat{\theta}^* - \mathbb{E}^*[\hat{\theta}^*])/\sigma_n$ для $\hat{\theta} = \text{med}(X_1, \dots, X_n)$ [$\sigma_n = (\mathbb{V}[\hat{\theta}])^{1/2}$]. Три графика слева: QQ-графики с целевыми значениями, обозначенными прямой. График справа: диаграммы размаха с целевым значением, обозначенными горизонтальной прямой. Всего 100 симуляций, 500 бутстраповских повторов для каждой симуляции.

На Рис. 2 показано качество бутстраповской аппроксимации для выборочной медианы в модели (18). AR-решетчатый бутстрап явно превосходит блочный бутстрап. Оценивание длины блока немного улучшает качество по сравнению с использованием фиксированной длины блока $\ell = 8 = n^{1/3}$: типичные значения $\hat{\ell}$ в 100 симуляциях равны 22, 16 и 30, что соответствует медианному значению и нижнему и верхнему квартилям. Наилучшее поведение AR-решетчатого бутстрапа неудивительно: в данном случае мы используем его преимущества,

обсужденные в подразделе 3.2. Данный результат указывает на количественный выигрыш в случае, когда истинный процесс не является AR-моделью конечного порядка и, следовательно, не является элементом аппроксимирующей решетки (для любого конечного размера выборки), но представим в виде процесса $AR(\infty)$ в (12). Как уже отмечено в Bühlmann (1997), преимущество AR-решетчатого бутстрапа обычно более значительно, если автоковариации процесса характеризуются некоторым затухающим псевдопериодическим убыванием, что верно для модели (18). Это свойство можно диагностировать графически, глядя на оцененные автоковариации.

Второй эксперимент проводился для нелинейного экспоненциального $AR(2)$ -процесса с гетероскедастичными инновациями:

$$\begin{aligned} X_t &= (0.5 + 0.9 \exp(-X_{t-1}^2))X_{t-1} - (0.8 - 1.8 \exp(-X_{t-1}^2))X_{t-2} + \sigma_t \epsilon_t, \\ \sigma_t^2 &= 0.5 + 0.1X_{t-1}^2 + 0.05\sigma_{t-1}^2 1_{[X_{t-1} \leq 0]} + 0.5 \exp(-\sigma_{t-1}^2) 1_{[X_{t-1} > 0]}, \end{aligned} \quad (19)$$

где $(\epsilon_t)_{t \in \mathbb{Z}}$ – IID-последовательность, независимая от $\{X_s; s < t\}$, $\epsilon_t \sim t_6/\sqrt{1.5}$. Этот процесс не представим в виде $AR(\infty)$ из (12).

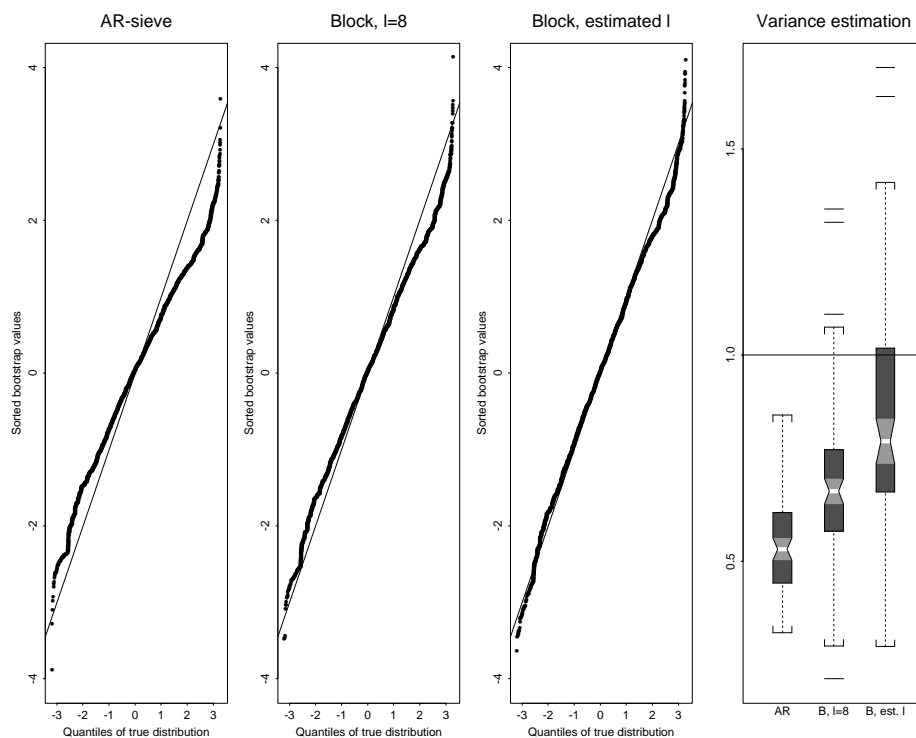


Рис. 3: Нелинейная модель (19), $n = 512$. Бутстраповское распределение и оценивание дисперсии $(\hat{\theta} - \mathbb{E}[\hat{\theta}])/\sigma_n$ с помощью $(\hat{\theta}^* - \mathbb{E}^*[\hat{\theta}^*])/\sigma_n$ для $\hat{\theta} = \text{med}(X_1, \dots, X_n)$ [$\sigma_n = (\mathbb{V}[\hat{\theta}])^{1/2}$]. Три графика слева: QQ-графики с целевыми значениями, обозначенными прямой. График справа: диаграммы размаха с целевым значением, обозначенными горизонтальной прямой. Всего 100 симуляций, 500 бутстраповских повторов для каждой симуляции.

На Рис. 3 показано качество бутстраповской аппроксимации для выборочной медианы в модели (19). AR-решетчатый бутстрап, являющийся асимптотически несостоятельным из-за нелинейности модели (19), дает явное смещение, и блочный бутстрап имеет превосходство. Как и в линейном случае (18) выше, использование оцененной длины блока $\hat{\ell}$ дает улучшение по сравнению с фиксированной длиной блока $\ell = 8 = n^{1/3}$: типичные значения $\hat{\ell}$ в 100 симуляциях равны 11, 7 и 14, что соответствует медианному значению и нижнему и верхнему квартилям. Результаты на Рис. 3 вновь дают количественное представление о выигрыше при использовании блочного бутстрапа в нелинейной модели.

4.2 Сравнение с методом подвыборок

Использование блочных подвыборок – это весьма общая техника оценивания моментов распределения оценки $\hat{\theta}$. Основная идея состоит в расчете оценки $\hat{\theta} = h_n(X_1, \dots, X_n)$ для большого числа подвыборок из ℓ последовательных наблюдений (блоков)

$$\hat{\theta}_{\ell,t} = h_\ell(X_{t-\ell+1}, \dots, X_t), \quad t = \ell, \dots, n.$$

Аппроксимации распределения и дисперсии с помощью подвыборок строятся следующим образом:

$$(n - \ell + 1)^{-1} \sum_{t=\ell}^n 1_{[a_\ell(\hat{\theta}_{\ell,t} - \hat{\theta}) \leq x]} \approx \mathbb{P}[a_n(\hat{\theta} - \theta) \leq x],$$

$$(n - \ell + 1)^{-1} \sum_{t=\ell}^n (a_\ell(\hat{\theta}_{\ell,t} - \hat{\theta}))^2 \approx a_n^2 \mathbb{V}(\hat{\theta}),$$

где $(a_n)_{n \in \mathbb{N}}$ – как в (2). Главным преимуществом данного метода по сравнению с бутстрапом

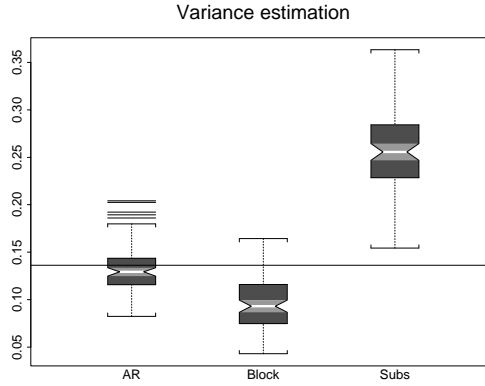


Рис. 4: Линейная модель (18), $n = 512$. Оценивание дисперсии $n\mathbb{V}[\hat{\theta}]$ для оцененной автокорреляции первого порядка $\hat{\theta} = \hat{\rho}(1)$ (горизонтальная прямая). AR, Block и Subs обозначают AR-решетчатый бутстрап, блочный бутстрап и метод подвыборок, соответственно; длина блоков для последних двух методов $\ell = 8$. Всего 100 симуляций, 500 бутстраповских повторов для каждой симуляции.

является весьма общие обстоятельства, в которых аппроксимация с помощью метода подвыборок является состоятельной. Детали описаны в Politis, Romano & Wolf (1999, раздел 3). Однако расчет оценки $\hat{\theta}$ в подвыборках гораздо меньшего размера $\ell \ll n$ и масштабирование ее поведения до размера исходной выборки n может быть проблематичным, если $\hat{\theta}$ сильно нелинейна, а размер выборки n не слишком велик. Простой, но впечатляющий пример – эмпирический аналог автокорреляции первого порядка

$$\hat{\theta} = \hat{\rho}(1) = \hat{R}(1)/\hat{R}(0),$$

как в примере А.

На Рис. 4 показаны результаты оценивания дисперсии $\hat{\rho}(1)$ в модели (18) при размере выборки $n = 512$. AR-решетчатый бутстрап дает наилучший результат, так как модель линейная; блочный бутстрап явно превосходит технику подвыборок при длине блоков $\ell = 8 = n^{1/3}$ для обоих методов (выбранной в соответствии с простым правилом выбора верной асимптотической скорости для обоих методов). В целом, не рекомендуется использовать технику подвыборок, когда известно, что бутстрап состоятелен. Техника подвыборок является интересным инструментом для сложных процедур с $\hat{\theta}$ (при большом объеме данных), когда бутстраповские схемы могут дать сбой.

5 Локальный бутстрап для оценок условного среднего

Решетчатый и блочный бутстрап дают разумные результаты для разнообразных оценок $\hat{\theta}$, если процесс, порождающий данные, принадлежит к подходящему для соответствующей бутстрап-схемы классу, как обсуждалось выше. Несколько удивительно, что некоторые схемы бутстрапа, основанные на независимом ресэмплинге, можно использовать для класса непараметрических оценок $\hat{\theta}$, имеющих более медленную скорость сходимости, чем $1/\sqrt{n}$, например, для $\hat{\theta}(\cdot)$ – (ядерной) сглаживающей оценки условного матожидания $\theta(\cdot) = \mathbb{E}[X_t | X_{t-1} = \cdot]$ стационарного процесса.

Сконцентрируемся на интервальном оценивании с помощью так называемого локального бутстрапа для условного матожидания $\theta(x) = \mathbb{E}[X_t | X_{t-1} = x]$ ($x \in \mathbb{R}$) стационарного процесса с областью действительных значений $(X_t)_{t \in \mathbb{Z}}$. Этот случай, выбранный ради простоты изложения, легко можно расширить на общий – $\mathbb{E}[f(X_t) | X_{t-i_1} = x_1, \dots, X_{t-i_p} = x_p]$ при заданном множестве p лагированных индексов $t - i_1, \dots, t - i_p$ и $f : \mathbb{R} \rightarrow \mathbb{R}$. Имея данные X_1, \dots, X_n , рассмотрим ядерную оценку

$$\hat{\theta}_h(x) = \frac{\sum_{t=2}^n W_{t,h}(x) X_t}{\sum_{t=2}^n W_{t,h}(x)}, \quad W_{t,h}(x) = K\left(\frac{x - X_{t-1}}{h}\right)$$

с шириной окна h .

Для бутстрапирования $\hat{\theta}_h(x)$ можно формировать псевдовыборки в рамках модели локальной регрессии

$$X_t^{*L} \sim \hat{F}_{X_{t-1},b} (t = 2, \dots, n) \text{ независимо от } X_s^{*L} (s \neq t),$$

где $\hat{F}_{x,b}(\cdot) = \sum_{t=2}^n W_{t,b}(x) \mathbb{I}_{[X_t \leq \cdot]} / \sum_{t=2}^n W_{t,b}(x)$ – оценка условной функции распределения X_t при $X_{t-1} = x$; b – это (контрольная) ширина окна, а веса $W_{t,b}(\cdot)$ определены выше. Тогда ресэмплинг осуществляется независимо от оцененной $\{\hat{F}_{x,b}(\cdot); x \in \mathbb{R}\}$, которой позволено меняться *локально*. Бутстрапированная ядерная оценка $\hat{\theta}_h(x)$ тогда получается из данных регрессионного типа $(X_1, X_2^{*L}), (X_2, X_3^{*L}), \dots, (X_{n-1}, X_n^{*L})$:

$$\hat{\theta}_h^{*L}(x) = \frac{\sum_{t=2}^n W_{t,h}(x) X_t^{*L}}{\sum_{t=2}^n W_{t,h}(x)}.$$

Такой подход был рассмотрен в Neumann & Kreiss (1998) и Paparoditis & Politis (2000).

Локальный бутстрап работает, поскольку асимптотическое распределение ядерной оценки $\hat{\theta}_h(x)$ является гауссовским, зависящим только от маргинального распределения X_t , условного распределения X_t на X_{t-1} и известной формы ядра; см. Robinson (1983). Локальный бутстрап способен состоятельно оценить все эти неизвестные. Верно это только в асимптотике, и для любого конечного размера выборки n уже дисперсия $\hat{\theta}_h(x)$ определенным образом зависит от n -мерного распределения $(X_t)_{t \in \mathbb{Z}}$. По построению, локальный бутстрап не может ухватывать зависимости за пределами условного распределения X_t на X_{t-1} . Этот недостаток отсутствует при использовании бутстрап-схем для зависимых наблюдений, например, блочного бутстрапа, точность оценивания $\mathbb{V}[\hat{\theta}_h(x)]$ выше при блочном бутстрапе, чем при локальном.

Neumann & Kreiss (1998) и Neumann (1998) строят (используя соответствующий локальный бутстрап) состоятельные доверительные области для $\theta(x)$, являющиеся *совместными* для x . Их скорости сходимости равны $1/\sqrt{n\bar{h}}$, как для поточечного случая. Это очень важный результат, так как аналитические совместные аппроксимации приближаются к предельному распределению экстремальных значений с очень медленной скоростью $1/\log(n)$: аналитический подход с предельным распределением гораздо хуже, чем схема локального бутстрапа.

5.1 Область применимости и выбор параметра настройки

Доказано, что локальный бутстрап состоятелен, если $(X_t)_{t \in \mathbb{Z}}$ – процесс с малым горизонтом зависимости, см. Paparoditis & Politis (2000) и Ango Nze, Bühlmann & Doukhan (2002).

Параметром настройки локального бутстрапа является контрольная ширина окна b . Простое решение – выбрать $b = h$, где h – предварительно выбранная ширина окна для оценивания $\hat{\theta}_h(x)$. Если b более высокого порядка, чем h , асимптотически нетривиальное смещение $\mathbb{E}[\hat{\theta}_h(x)] - \theta(x)$ можно оценить с помощью локального бутстрапа, см. Paparoditis & Politis (2000). Контрольная ширина окна играет роль при оценивании условного распределения X_t на X_{t-1} : эта задача довольно проста для двумерных распределений. Процедура не является чувствительной к спецификации этой контрольной ширины окна.

5.2 Локальный и блочный бутстрап для симулированных временных рядов

Интересно узнать, имеет ли преимущество, с точки зрения конечных выборок, бутстрап-схема, принимающая во внимание временную структуру, например блочный бутстрап. Рассмотрим

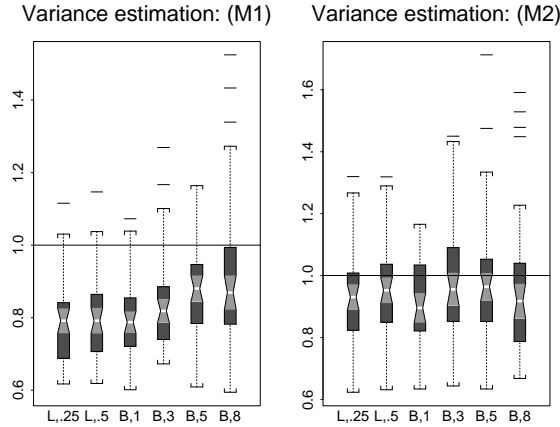


Рис. 5: Бутстраповские оценки дисперсии $\mathbb{V}^*[\hat{\theta}_h^*(x)]/\mathbb{V}[\hat{\theta}_h(x)]$ при $x = 1, 60$ и $x = 0, 76$ для (M1) и (M2) соответственно (целевой уровень указан горизонтальной прямой); L соответствует локальному бутстрапу с контрольной шириной окна 0,25 и 0,5; B соответствует блочному бутстрапу с длиной блока 1, 3, 5 и 8. Размер выборки $n = 512$.

симуляционный эксперимент на основе билинейной модели:

$$X_t = 0.5\epsilon_{t-1}X_{t-1} + \epsilon_t, \quad (20)$$

где $(\epsilon_t)_{t \in \mathbb{Z}}$ – последовательность IID-инноваций, где ϵ_t независимы от $\{X_s; s < t\}$. Рассмотрим случаи, когда

$$(M1) \quad \epsilon_t \text{ IID} \sim U(\{-1, 1\}), \text{ то есть } \mathbb{P}[\epsilon_t = 1] = \mathbb{P}[\epsilon_t = -1] = 1/2,$$

$$(M2) \quad \epsilon_t \text{ IID} \sim U([-1, 1]).$$

Обе модели (M1) и (M2) характеризуются слабой формой зависимости. Эмпирически мы обнаружили, что оценивание $\mathbb{V}[\hat{\theta}_h(x)]$ сложнее для модели с дискретными инновациями (M1), чем для (M2).

Рис. 5 отражает результаты оценивания $\mathbb{V}[\hat{\theta}_h(x)]$ со стандартным гауссовским ядром K и разумной шириной окна $h = 0, 25$. Размер выборки $n = 512$. Графически определить существенные различия затруднительно. Количественное описание имеет следующий вид. Для модели (M1) блочный бутстрап с $\ell = 5$ ('B,5') дает наилучшие результаты в смысле MSE:

он дает приблизительно на 40% меньшую MSE, чем наилучший вариант локального бутстрапа с $b = 0,5$ ('L,.5'): двусторонний парный тест Уилкоксона отдает предпочтение схеме 'B,5' с P -значением 0,002 при нулевой гипотезе о равных MSE. Сравнение любого варианта локального бутстрапа с любым другим вариантом блочного бутстрапа с длинами блоков $\ell = 1, 3, 8$ не дает значительных отличий. Для (M2) блочный бутстрап с $\ell = 1$ (что является регрессионным бутстрапом при независимости) дает наилучшие результаты с MSE, на приблизительно 9% меньшей, чем для наилучшего локального бутстрапа с $b = 0,5$, но различие незначимо. За исключением варианта 'B,8' с неразумно большой длиной блоков (значительно уступающего локальному бутстрапу), сравнение любого варианта локального бутстрапа с любыми вариантами блочного бутстрапа при $\ell = 1, 3, 5$ не выявляет значимого различия. Для конкретной билинейной модели со слабой степенью зависимости можно сделать следующий вывод. В более простом случае (M2) локальный и блочный бутстрап одинаково хороши (исключая случай неразумной длины блоков $\ell = 8$). Это не так для немного более сложного случая (M1), когда блочный бутстрап всегда так же хорош, как локальный, или даже лучше него, при условии, что имеется удачное правило выбора длины блока вблизи $\ell = 3$. Ожидается более явное преимущество блочного бутстрапа по сравнению с локальным, если данные характеризуются большей степенью зависимости.

6 Заключение

Среди схем блочного, решетчатого и локального бутстрапа блочный бутстрап является наиболее общим методом. Другое его преимущество – простая реализация процедуры ресэмплинга, не сложнее, чем в эфроновском бутстрапе для независимых наблюдений. Недостатки же метода заключаются в следующем. Блочно-бутстраповскую выборку не следует рассматривать как разумную выборку, имитирующую порождающий данные процесс: она нестационарна и искажена в местах соединения взятых из выборки блоков. Это означает, что правило подстановки для бутстрапирования оценки $\hat{\theta}$ не годится. Предварительная векторизация данных очень рекомендуется, но это может потребовать изменения бутстрапированной оценки и процедуры ее подсчета. Как общая непараметрическая схема, блочный бутстрап может уступать другим методам для различных классов стационарных временных рядов, например, для линейных. Точность второго порядка для доверительного интервала достигается путем стьюдентизации и BC_a -коррекции; последний метод дает маргинальное улучшение в симулированном примере. Двойное бутстрапирование не дает больших надежд, так как блочный бутстрап на первой итерации нарушает зависимость в местах соединения блоков.

Решетчатый бутстрап в целом отталкивается от разумной модели временного ряда. Это приводит к двум преимуществам: правило подстановки используется для определения и расчета бутстраповской оценки, а двойной бутстрап потенциально приводит к точности более высокого порядка. Хорошие схемы решетчатого бутстрапа, такие как AR-решетчатый бутстрап, адаптируются к степени зависимости временного ряда: их точность растет по мере снижения степени зависимости, см. формулу (15). Это не так для блочного бутстрапа, как видно из формулы (8). Кроме того, решетчатый бутстрап в целом менее чувствителен к выбору модели для решетки, чем блочный бутстрап к длине блоков.

AR-решетчатый бутстрап, естественно, является наилучшим, если процесс, порождающий данные, является линейным временным рядом, представимым в виде $AR(\infty)$, см. (12). Метод легок в реализации благодаря простоте подгонки AR-модели.

Локальный бутстрап из раздела 5 ограничивается процедурами непараметрического оценивания, имеющими более низкую скорость сходимости, чем $1/\sqrt{n}$. Хотя он реализуется аналогично регрессионному бутстрапу в случае независимости, он состоятелен и, следовательно, устойчив к некоторым формам зависимости. Его преимуществом является простота, поскольку не требуется специфицировать параметр настройки, управляющий степенью за-

висимости порождающего данные процесса. С другой стороны, это также свидетельствует о его слабости и неспособности имитировать зависимость подобающим образом: этот метод может давать результаты похуже, чем блочный бутстрап.

7 Другие результаты и примечания к списку литературы

В дополнение к приведенному избирательному изложению дадим ссылки на дополнительную литературу. Efron & Tibshirani (1993, гл. 8.5–8.6), Shao & Tu (1995, гл. 9), Li & Maddala (1996), Davison & Hinkley (1997, гл. 8) обсуждают методы бутстрапирования для зависимых наблюдений с другой точки зрения, чем наш сравнительный обзор.

Литература по блочному бутстрапу на настоящий момент обширна. Обзор более ранних исследований в этой области содержится в Léger, Politis & Romano (1992). Переработка результатов Künsch (1989) при минимальных предположениях содержится в Radulović (1996). Различные результаты для эмпирических процессов приводятся в Bühlmann (1994, 1995) и Peligrad (1998). Lahiri (1996) доказывает корректность второго порядка для блочного бутстрапа в случае, когда $\hat{\theta}$ является М-оценкой в линейной регрессионной модели с зависимым шумом. Техника блочного бутстрапа также применима к пространственным процессам, см. Politis & Romano (1993). Версия блочного бутстрапа, достигающая стационарности бутстраповской выборки (так называемый стационарный бутстрап), была предложена в Politis & Romano (1994a). Lahiri (1999) строго показывает, что блочный бутстрап лучше стационарного. Carlstein, Do, Hall, Hesterberg & Künsch (1998) предлагают схему связывания блоков при формировании псевдovyборки: они утверждают, что для случая оценивания дисперсии $\hat{\theta} = \overline{X}_n$ такая процедура имеет меньшую среднеквадратическую ошибку.

Связанными с блочным бутстрапом являются методы, использующие подвыборки. Работу Carlstein (1986) можно рассматривать как предшественницу блочного бутстрапа для оценивания дисперсии. В своей выдающейся статье Politis & Romano (1994b) показали, что метод подвыборок применим гораздо в более общем случае, чем методы блочного бутстрапа, а именно, практически во всех случаях, когда $\hat{\theta}$ имеет некоторое невырожденное предельное распределение. Künsch (1989) утверждает, что для случая, когда статистика $\hat{\theta}$ асимптотически нормальна, блочный бутстрап превосходит метод подвыборок. Другие результаты в методе подвыборок можно найти в книге Politis, Romano & Wolf (1999).

Основанное на моделях бутстрапирование изучалось для множества случаев: Freedman (1984) – для AR-модели; Kreiss & Franke (1992) – для ARMA-моделей; Paparoditis & Politis (2002) – для марковских моделей. Непараметрическая AR(1)-модель с гетероскедастичными инновациями обсуждается в Franke, Kreiss, Mammen & Neumann (2002): бутстрап, основанный на моделях, можно использовать для точного построения совместных доверительных интервалов для авторегрессионной функции $m(x) = \mathbb{E}[X_t | X_{t-1} = x]$. Отметим, что того же можно достигнуть (в первом порядке) с помощью локального бутстрапа из раздела 5.

Для AR-решетчатого бутстрапа результаты для эмпирических процессов даны в Bickel & Bühlmann (1999) путем введения слабого понятия перемешивания для бутстрапируемых процессов. Нестационарный случай $X_t = m_t + Z_t$ ($t \in \mathbb{Z}$), где $(m_t)_{t \in \mathbb{Z}}$ – медленно меняющийся детерминистский тренд, а $(Z_t)_{t \in \mathbb{Z}}$ – AR(∞)-шум, изучается в Bühlmann (1998), где обсуждаются AR-решетчато-бутстраповские доверительные интервалы для тренда.

Объединение решетчатых или модельных методов с блочным бутстрапом было предложено в Davison & Hinkley (1997, гл. 8.2); авторы называют такую процедуру ‘последующее зачернение’. Идея состоит в том, чтобы предварительно отбелить временной ряд с помощью модельного или решетчатого подхода, а затем применить блочный бутстрап к, возможно, менее зависимым, отбеленным остаткам: блочная формировка подвыборок из этих остатков и обращение операции отбеливания тогда дают зачерненную бутстраповскую выборку.

Другой метод бутстрапирования стационарных линейных временных рядов был предложен

в Dahlhaus & Janas (1996); авторы независимым образом ресэмплируют значения периодограммы в частотной области соответственно оценке спектральной плотности. По построению, такой ресэмплинг затрагивает только структуру автоковариаций, и состоятельность таким образом ограничивается только линейными временными рядами. Идея независимого ресэмплинга в частотной области прежде появлялась в Franke & Härdle (1992) при бутстрапировании оценки спектральной плотности; ее модификация для бутстрап-схемы локального типа рассматривается в Paparoditis & Politis (1999).

Список литературы

- Ango Nze, P., P. Bühlmann & P. Doukhan (2002). Weak dependence beyond mixing and asymptotics for nonparametric regression. *Annals of Statistics* 30, 397–430.
- Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* 74, 457–468.
- Bickel, P.J. & P. Bühlmann (1997). Closure of linear processes. *Journal of Theoretical Probability* 10, 445–479.
- Bickel, P.J. & P. Bühlmann (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* 5, 413–446.
- Bühlmann, P. (1994). Blockwise bootstrapped empirical process for stationary sequences. *Annals of Statistics* 22, 995–1012.
- Bühlmann, P. (1995). The blockwise bootstrap for general empirical processes of stationary sequences. *Stochastic Processes and their Applications* 58, 247–265.
- Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* 3, 123–148.
- Bühlmann, P. (1998). Sieve bootstrap for smoothing in non-stationary time series. *Annals of Statistics* 26, 48–83.
- Bühlmann, P. (2002). Bootstraps for time series. *Statistical Science* 17, 52–72.
- Bühlmann, P. & H.R. Künsch (1995). The blockwise bootstrap for general parameters of a stationary time series. *Scandinavian Journal of Statistics* 22, 35–54.
- Bühlmann, P. & H.R. Künsch (1999). Block length selection in the bootstrap for time series. *Computational Statistics and Data Analysis* 31, 295–310.
- Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics* 14, 1171–1179.
- Carlstein, E., K.-A. Do, P. Hall, T. Hesterberg & H.R. Künsch (1998). Matched-block bootstrap for dependent data. *Bernoulli* 4, 305–328.
- Choi, E. & P. Hall (2000). Bootstrap confidence regions computed from autoregressions of arbitrary order. *Journal of Royal Statistical Society (Series B)* 62, 461–477.
- Dahlhaus, R. & D. Janas (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *Annals of Statistics* 24, 1934–1963.
- Davison, A.C. & P. Hall (1993). On studentizing and blocking methods for implementing the bootstrap with dependent data. *Australian Journal of Statistics* 35, 215–224.
- 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. (1987). Better bootstrap confidence intervals. *Journal of American Statistical Association* 82, 171–185.
- Efron, B. & R.J. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman & Hall.
- Franke, J. & W. Härdle (1992). On bootstrapping kernel spectral estimates. *Annals of Statistics* 20, 121–145.
- Franke, J., J.-P. Kreiss, E. Mammen & M.H. Neumann (2002). Properties of the nonparametric autoregressive bootstrap. *Journal of Time Series Analysis* 23, 555–585.
- Freedman, D.A. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *Annals of Statistics* 12, 827–842.
- Götze, F. & H.R. Künsch, (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics* 24, 1914–1933.

- Grenander, U. (1981). *Abstract Inference*. Wiley, New York.
- Hall, P. (1985). Resampling a coverage pattern. *Stochastic Processes and their Applications* 20, 231–246.
- Hall, P. (1986). On the bootstrap and confidence intervals. *Annals of Statistics* 14, 1431–1452.
- Hall, P., J.L. Horowitz & B.-Y. Jing (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* 82, 561–574.
- Hall, P., B.-Y. Jing & S.N. Lahiri (1998). On the sampling window method under long range dependence. *Statistica Sinica* 8, 1189–1204.
- Kreiss, J.-P. (1992). Bootstrap procedures for $AR(\infty)$ -processes. В *Bootstrapping and Related Techniques*, Eds. Jöckel, K.-H., G. Rothe & W. Sendler, стр. 107–113. Springer, Heidelberg.
- Kreiss, J.-P. & J. Franke (1992). Bootstrapping stationary autoregressive moving-average models. *Journal of Time Series Analysis* 13, 297–317.
- Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.
- Lahiri, S.N. (1993). On the moving block bootstrap under long range dependence. *Statistics and Probability Letters* 18, 405–413.
- Lahiri, S.N. (1995). On the asymptotic behaviour of the moving block bootstrap for normalized sums of heavy-tail random variables. *Annals of Statistics* 23, 1331–1349.
- Lahiri, S.N. (1996). On Edgeworth expansion and moving block bootstrap for studentized M-estimators in multiple linear regression models. *Journal of Multivariate Analysis* 56, 42–59.
- Lahiri, S.N. (1999). Theoretical comparisons of block bootstrap methods. *Annals of Statistics* 27, 386–404.
- Léger, C., D.N. Politis & J.P. Romano (1992). Bootstrap technology and applications. *Technometrics* 34, 378–398.
- Li, H. & G.S. Maddala (1996). Bootstrapping time series models. *Econometric Reviews* 15, 115–158.
- Loh, W. (1987). Calibrating confidence coefficients. *Journal of American Statistical Association* 82, 155–162.
- Martin, R.D. & V.J. Yohai (1986). Influence functionals for time series. *Annals of Statistics* 14, 781–818 (обсуждение на стр. 819–855).
- Neumann, M.H. (1998). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Annals of Statistics* 26, 2014–2048.
- Neumann, M.H., J.-P. Kreiss (1998). Regression-type inference in nonparametric autoregression. *Annals of Statistics* 26, 1570–1613.
- Paparoditis, E. & D.N. Politis (1999). The local bootstrap for periodogram statistics. *Journal of Time Series Analysis* 20, 193–222.
- Paparoditis, E. & D.N. Politis (2000). The local bootstrap for kernel estimators under general dependence conditions. *Annals of the Institute of Statistical Mathematics* 52, 139–159.
- Paparoditis, E. & D.N. Politis (2001). Tapered block bootstrap. *Biometrika* 88, 1105–1119.
- Paparoditis, E. & D.N. Politis (2002). The local bootstrap for Markov processes. *Journal of Statistical Planning and Inference* 108, 301–328.
- Peligrad, M. (1998). On the blockwise bootstrap for empirical processes for stationary sequences. *Annals of Probability* 26, 877–901.
- Politis, D.N. & J.P. Romano (1992). A general resampling scheme for triangular arrays of α -mixing random variables. *Annals of Statistics* 20, 1985–2007.
- Politis, D.N. & J.P. Romano (1993). Nonparametric resampling for homogeneous strong mixing random fields. *Journal of Multivariate Analysis* 47, 301–328.
- Politis, D.N. & J.P. Romano (1994a). The stationary bootstrap. *Journal of American Statistical Association* 89, 1303–1313.
- Politis, D.N. & J.P. Romano (1994b). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics* 22, 2031–2050.
- Politis, D.N., J.P. Romano & M. Wolf (1999). *Subsampling*. Springer.

- Radulović, D. (1996). The bootstrap of the mean for strong mixing sequences under minimal conditions. *Statistics and Probability Letters* 28, 65–72.
- Robinson, P.M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–207.
- Shao, J. & D. Tu (1995). *The Jackknife and Bootstrap*. Springer.
- Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics* 8, 147–164.

Bootstrap schemes for time series

Peter Bühlmann

ETH Zürich, Zürich, Switzerland

We review and compare block, sieve and local bootstraps for time series and thereby illuminate theoretical aspects of the procedures as well as their performance on finite-sample data. Our view is selective with the intention of providing a new and fair picture of some particular aspects of bootstrapping time series. The generality of the block bootstrap is contrasted with the sieve bootstrap. We discuss implementational advantages and disadvantages, and argue that the sieve often outperforms the block method. Local bootstraps, designed for nonparametric smoothing problems, are easy to use and implement but exhibit in some cases low performance.