

Материалы для лекций по временным рядам (Голяндина Н.Э.)

Оглавление

Глава 1. Введение	3
1.1. Первый тип рядов: тренд + периодическая компонента + шум = сигнал + шум.	3
1.2. Второй тип рядов: ряд — случайный процесс	6
1.3. О чем пойдет речь в данном курсе	7
Глава 2. Лекция 2. Частотный анализ. Периодограмма	8
2.1. Периодограмма временного ряда	9
2.2. Периодическая компонента	10
2.2.1. Одна гармоника	10
2.2.2. Произвольная периодическая компонента	11
2.3. Тренд: периодограмма тренда	12
2.4. Шум и его периодограмма	13
Глава 3. Лекция 3. Линейные фильтры	16
3.1. Линейный фильтр, импульсная характеристика, причинный фильтр, FIR	16
3.2. Характеристики фильтра через его воздействие на $\cos(2\pi\omega n)$ (или на комплексную экспоненту) — АЧХ, ФЧХ	17
3.3. АЧХ фильтра скользящего среднего, зависимость от длины окна	18
3.4. АЧХ фильтра перехода к разностям (дифференцирования)	19
3.5. Связь между периодограммами ряда до и после применения фильтра	20
3.6. Что такое запаздывание и отчего оно может возникать (на примере скользящего среднего)?	22
3.7. Фильтр как оценка сигнала	24
3.7.1. Дисперсия	24
3.7.2. Смещение, роль второй производной	25

Глава 1

Введение

Пусть $X = (x_1, \dots, x_N)$ — последовательность измерений, ряд длины N . Обычно n — это время, а $x_i = f(\delta i)$ — измерения некоторой функции с равным шагом. Часто измерения содержат случайные ошибки, называемые шумом (позже сформулируем, как формализуется шум).

Примеры с картинками: ...

Задача: выявить структуру во временном ряде и ее спрогнозировать. Давайте разберемся, что это за структура.

Важно понимать, что нельзя просто прогнозировать временной ряд, можно повторять в будущем только что-то, что было обнаружено в настоящем. Далее мы коротко коснемся таких закономерностей.

Замечание. В принципе, не всегда ставят задачу прогноза. Например, ставят задачу выделения тренда (тенденции). Но зачем? Очень часто — чтобы мысленно продолжить ее в будущее.

Задача анализа временного ряда тоже можно быть важна сама по себе, в том числе, с исторической точки зрения. Задачу анализа временного ряда можно пробовать рассматривать как чисто математическую. К сожалению, задача прогнозирования не является математической и невозможно гарантировать какие-то свойства прогноза, так как продолжаемая в будущем структура может меняться в силу внешних обстоятельств.

1.1. Первый тип рядов: тренд + периодическая компонента + шум = сигнал + шум.

$$x_n = t_n + p_n + \epsilon_n = s_n + \epsilon_n, \mathbb{E}\epsilon_n = 0.$$

Сигнал — неслучайная (детерминированная) составляющая ряда, в нем и содержится структура для такого типа рядов.

Тренд Нет одного формального определения:

самое общее — медленно меняющаяся составляющая временного ряда

параметрическое: аппроксимация ряда из какого-то параметрического семейства, например, линейные функции или полиномы порядка не больше, чем заданный. Терминология: линейный тренд, полиномиальный тренд.

Параметрические виды тренда: линейный $t_n = an + b$, квадратичный $t_n = an^2 + bn + c$, экспоненциальный $t_n = C \exp \alpha n$.

Замечание: $\exp \alpha n = \rho^n$ при $\rho = \exp \alpha$. Поэтому функцию $C\rho^n$ тоже называют экспоненциальной. Обычно $C > 0$. Тогда, если $\alpha > 0$ ($\rho > 1$), то это возрастающая экспонента, если там равенство, то это константа; иначе — убывающая к нулю экспонента.

Примеры с картинками.

То, что нет точного определения, объективно. В частности, понятие медленных изменений явно зависит от той части ряда, которую наблюдаем. Например, синус для большой промежутке — явно не тренд, а на промежутке типа четверть периода не отличим от тренда.

Имеет смысл отличать от тенденции: тенденция обычно монотонно возрастает или убывает.

Часто в стат.пакетах под словом тренд понимают линейный тренд, а тренд в общем смысле представляют в виде тренда-тенденции и циклических (медленных нерегулярных колебаний).

Проблема с трендами: как правило, они определяются внешними условиями, т.е. плохо предсказуемы. Есть редкие случаи, когда тренд, например, экспоненциально растет (пример...), даже в этом случае экспоненциальный рост может прекратиться. Например, закон Мальтуса, когда количество людей растет по экспоненте, а количество еды/ресурсов — как линейная функция. Проблема! (хотя не знаю, как сейчас выглядит поведение числа людей и количества ресурсов)

Или: все было хорошо, издали плохой закон, все стало плохо (это условный пример, просто идея).

Или финансовый индекс S&P рос экспоненциально, пока не случился кризис.

Иногда говорят о случайных трендах. Но это, скорее, относится уже к следующему подразделу.

Периодическая компонента Периодическая компонента с периодом T : то, что повторяется каждые T точек: $p_{n+T} = p_n$ для любого n .

Пример: синус с периодом T . Пишем формулу.

$$p_n = A \cos(2\pi n/T + \phi) = A \cos(2\pi \omega n + \phi).$$

Запишите как сумму синуса и косинуса по формуле косинус суммы равен произведению косинусов минус произведение синусов. (Это потом пригодится в разложении Фурье.)

Частота $\omega = 1/T$. (Циклическая частота — $2\pi/T$, но мы ее не будем использовать.)

ϕ — фаза. С помощью фазы можно синус превращать в косинус и наоборот. Поэтому можно называть компоненту $A \cos(2\pi \omega n + \phi)$ гармоникой.

A — амплитуда. Если амплитуда не константа, а медленно меняется (опять это неформальное выражение; имеется в виду, что существенно медленнее, чем сам синус/косинус), то гармоника называется модулированной. Например, $p_n = \exp \alpha n A \cos(2\pi \omega n + \phi)$ — экспоненциально-модулированная гармоника.

Для временного ряда период может быть только целым, для косинуса/синуса — не обязательно.

Как, например, можно составить период 12 из шести косинусов (гармоник) с разными периодами?

Ответ: сумма косинусов с периодами 12, $6 = 12/2$, $4 = 12/3$, $3 = 12/4$, $2.4 = 12/5$, $2 = 12/6$.

Относительность понятия тренда: можно нарисовать растущую половину синуса, спросить, похоже ли на тренд (ответ — да), а потом дорисовать этот синус.

Периодичности, как правило, порождены внутренними свойствами временного ряда (мороженое чаще покупают летом), поэтому предсказывать периодичность — гораздо более благодарная задача, чем предсказывать тренд.

Часто предсказывать нужно именно тренд/тенденцию. Тогда периодичность является мешающим фактором и ее удаляют (иначе, например, легко спутать сезонное уменьшение продаж с просто падением продаж (или наоборот)). Экономические временные ряды часто выкладываются в виде *seasonally adjusted*, когда из ряда удаляется сезонная периодичность (шум остается).

Сигнал Если нет периодичности, то сигнал = тренд.

Без разделения на тренд и периодичность, сигнал можно определить как мат.ож. временного ряда, так как мат.ож. шума равно нулю, а сигнал не случайный.

Шум Шум — в данной модели можно понимать как нерегулярные быстрые колебания вокруг нуля. Формально - это набор случайных величин с нулевым математическим ожиданием. Часто модель — стационарный случайный процесс.

В данной модели шум — мешающая компонента ряда, его нужно устранять, подавлять и пр. В след. подразделе рассматривается структура случайного процесса, которая позволяет прогнозировать и его.

Методы Методы выделения тренда, методы разложения на компоненты, сглаживание, фильтры. Прогноз тренда и периодик.

1.2. Второй тип рядов: ряд — случайный процесс

Случайный процесс (с дискретным временем) ξ_1, \dots, ξ_N состоит из последовательности случайных величин. Т.е. это многомерный случайный вектор. Самый стандартный класс случайных процессов — стационарные случайные процессы. Стационарные процессы в слабом смысле: ξ_k имеют одинаковое математическое ожидание (часто равно нулю), одинаковую дисперсию и корреляция/ковариация зависит только от разницы во времени: $\text{Cov}(\xi_t, \xi_s) = V(t - s)$.

Если значения ряда — случайные величины, то и прогноз (продолжение ряда) — тоже случайная величина или некое распределение, которое задается плотностью или вероятностями, с которыми случаются те или иные исходы.

В этом случае точкой прогноза считается математическое ожидание этого распределения. Но более правильно строить предсказательный интервал, который показывает, какие значения могут случиться с заданной вероятностью, например, 0.95.

Например, $x_n = \epsilon_n$, $\mathbb{E}\epsilon_n = 0$, ϵ_n одинаково распределены и независимы между собой (iid — independent, identically distributed). Это белый шум. Структура? Предсказание? Увы, структуры нет, среднее предсказания равно нулю. Но если оценить разброс шума (стандартное отклонение s), то 95% доверительный интервал — $(\pm 1.96s)$. Даже такая информация полезна. Если что-то в регулярном режиме колеблется как белый шум, то при выходе из предсказательного интервала несколькими значениями подряд можно делать вывод о том, что что-то сломалось.

Другой пример: $x_n = \phi x_{n-1} + \sigma \epsilon_n$. Вопрос: если $\sigma = 0$, то как ведет себя ряд при разных ϕ ? Ответ: получили экспоненциальный ряд, который стремится к нулю при $\phi < 1$. Это не случайный и не стационарный ряд.

Оказывается, что если $|\phi| < 1$ и добавить белый шум на каждом шаге, то получим стационарный случайный процесс. Если ϕ еще и больше 0, то этот процесс называется красным шумом. За счет чего такие процессы можно прогнозировать? За счет корреляции: $\rho(\xi_n, \xi_{n+1}) = \phi$.

А что будет, если $\phi = 1$? Это называется броуновским движением: $x_n = x_{n-1} + \epsilon_n$. Это НЕстационарный случайный процесс. У него растет стандартное отклонение (корень из дисперсии) как \sqrt{n} . Однако, если перейти к последовательным разностям: $y_n = x_{n+1} - x_n$, то получим стационарный процесс (белый шум). В этом случае говорят, что ряд имел случайный тренд.

Методы Авторегрессия (ARIMA, Seasonal ARIMA), экспоненциальное сглаживание (ETS).

1.3. О чем пойдет речь в данном курсе

TODO

Глава 2

Лекция 2. Частотный анализ. Периодограмма

Вообще, всегда интересуют изменения во времени. Собственно, понимание этих изменений — это и есть анализ временных рядов.

Изменения не обязательно регулярные и их сложно описывать. Гораздо удобнее свести нерегулярные колебания к сумме регулярных.

А что такое регулярные колебания? Самые простые — это так называемые гармоники, т.е. синусы и косинусы.

В общем виде это косинус с периодом T . $h_n = A \cos(2\pi n/T + \phi) = A \cos(2\pi\omega n + \phi)$, $n = 1, \dots, N$.

Терминология: T — период, ω — частота, A — амплитуда, ϕ — фаза. С помощью фазы можно синус превращать в косинус и наоборот: $\sin(x + \pi/2) = \cos(x)$. Поэтому часто называют такую компоненту $A \cos(2\pi\omega n + \phi)$ гармоникой.

Почему тут T — это период?

Период — понятно, что такое, это число точек, через которое значения временного ряда повторяются.

Частота $\omega = 1/T$ — обратная величина. Чем больше период, тем реже точки повторяются. И наоборот.

Для гармоник понятие частоты можно обобщить на случай, когда $T = 1/\omega$ не обязательно целое. Будем это использовать.

Однако, на частоту нужно наложить естественные ограничения: $0 \leq \omega \leq 0.5$. 0 соответствует особому случаю, когда ряд — константа, так как $p_n = \cos(0) = 1$. 0.5 соответствует другому особому случаю, когда ряда — ‘пила’, так как $p_n = \cos(\pi n) = (-1)^n$.

Колебаться чаще, чем с частотой 0.5 (период 2) ряд не может.

Можно рассмотреть еще более простые элементы — синусы и косинусы без фазы, воспользовавшись соотношением $A \cos(2\pi\omega n + \phi) = C \cos(2\pi n\omega) + S \sin(2\pi n\omega)$, где $A^2 = C^2 + S^2$, $\phi = \arctan(\dots)$.

Итак, рассмотрим разложение Фурье, которое сводит произвольные колебания к сумме гармонических.

2.1. Периодограмма временного ряда

Пусть $X = (x_1, \dots, x_N)$ — временной ряд длины N . Рассмотрим ортогональный тригонометрический базис, состоящий из векторов $(\cos 2\pi n \frac{k}{N}, n = 1, \dots, N)^T$ и $(\sin 2\pi n \frac{k}{N}, n = 1, \dots, N)^T$ при $k = 0, \dots, \lfloor \frac{N}{2} \rfloor$. При $k = 0$ вектора с $\sin(0)$ нет, он нулевой, а $\cos(0) = 1$. А если N четное, то нет и вектора с $\sin(\pi n) = 0$ при $k = N/2$, а $\cos(\pi n) = (-1)^n$.

Размер базиса равен N . Для $N = 2m + 1$ получим размер $1 + 2\lfloor \frac{2m+1}{2} \rfloor = 1 + 2m = N$. Для $N = 2m$ получим $1 + 2\lfloor \frac{2m}{2} \rfloor = 1 + 2m = N + 1$, но последний синус будет нулевым, поэтому размер базиса равен N .

Итак, мы хотим получить представление

$$x_n = C_0 + \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \{C_k \cdot \cos\left(2\pi n \frac{k}{N}\right) + S_k \cdot \sin\left(2\pi n \frac{k}{N}\right)\},$$

где $n = 1, \dots, N$. Это называется *разложением Фурье* по тригонометрическому базису. Эквивалентное представление разложения Фурье:

$$x_n = A_0 + \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} A_k \cos\left(2\pi n \frac{k}{N} + \phi_k\right),$$

где $A_k^2 = C_k^2 + S_k^2$, $A_0 = c_0$.

Понятно, что по коэффициентам A_k можно понять, как ведет себя ряд.

Можно ввести функцию от частоты, заданную на решетке, она называется *периодограммой ряда*:

$$\Pi\left(\frac{k}{N}\right) = \frac{N}{2} \begin{cases} 2A_0^2 & \text{if } k = 0 \\ 2A_{\frac{N}{2}}^2 & \text{if } N \text{ is even and } k = \frac{N}{2} \\ A_k^2 & \text{otherwise} \end{cases}$$

Объяснение того, что на крайних частотах коэффициент другой, состоит в том, что квадраты норм этих косинусов в два раза больше.

У периодограммы временного ряда есть разные нормировки. В частности, часто нет коэффициента N . В данном случае нормировка такая, чтобы квадрат евклидовой нормы временного ряда был равен сумме значений периодограммы.

На графиках частот на N не умножают. Тогда сумма значений периодограммы будет равна средней сумме квадратов значений ряда. Мы будем рассматривать именно такой вариант.

Давайте научимся ‘читать’ периодограммы на основе ее графика. Сначала будем по ряду смотреть, какая у него периодограмма. Разобрав это, сможем наоборот по периодограмме понимать, какой ряд.

2.2. Периодическая компонента

2.2.1. Одна гармоника

Пусть $x_n = A \cos(2\pi\omega n + \phi)$, $0 < \omega \leq 0.5$.

Казалось бы, в разложении Фурье только одно ненулевое слагаемое, но (!) это так только если частота попадает в решетку k/N ($N\omega$ — целое; по другому — длина ряда кратна периоду). Тогда периодограмма выглядит так:

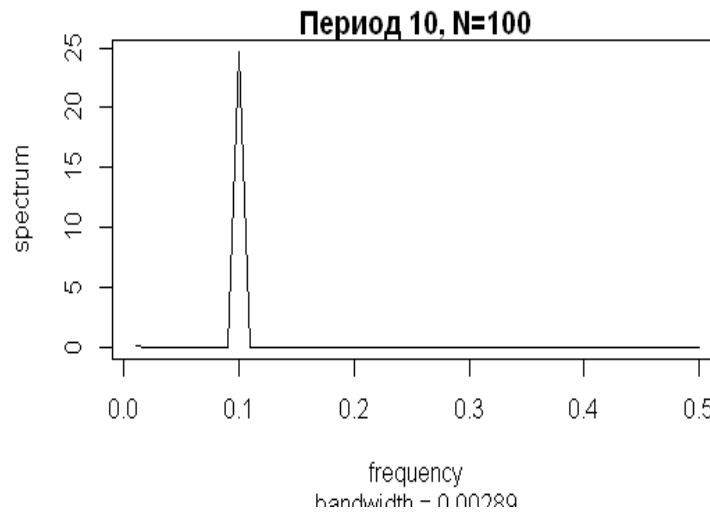


Рис. 2.1. Периодограмма без растекания частоты

Имеем одно ненулевое значение в точке $\omega = k_0/N$, значение периодограммы A^2 .

Если длина ряда не будет кратна периоду, то получим так называемое растекание частоты (frequency leakage).

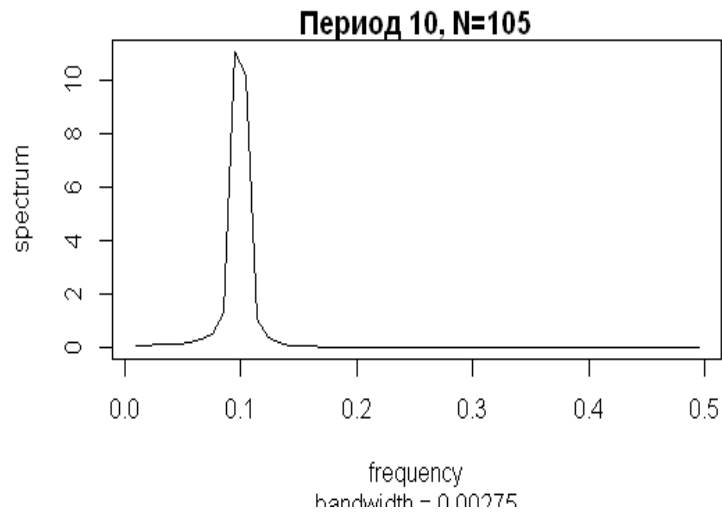


Рис. 2.2. Периодограмма исходного ряда

Поэтому, если есть возможность, то либо добавляют нули к ряду, что длина ряда делилась на период, либо удаляют точки из начала ряда (лучше из начала, так как последних точек жалко). Конечно, так можно сделать, только если период известен (например, сезонность).

2.2.2. Произвольная периодическая компонента

Можно показать, что произвольная периодическая компонента с (целым) периодом T является суммой гармоник с частотами k/T (если N кратно периоду), так как ее разложение Фурье сводится к разложению Фурье одного периода длины T .

Это эквивалентно тому, что на периодограмме будут (вернее, могут быть) ненулевые значения в точках k/T .

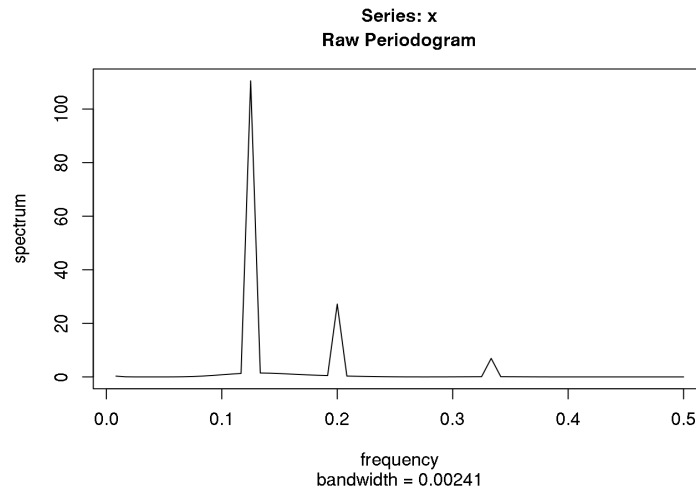


Рис. 2.3. Периодограмма периодического ряда

Пример: данные по месяцам и сезонное поведение. Получим сумму косинусов с периодами 12, $6 = 12/2$, $4 = 12/3$, $3 = 12/4$, $2.4 = 12/5$, $2 = 12/6$.

Конечно, если длина ряда не кратна периоду, то мы получим растекание вокруг каждой из частот.

Замечание: для амплитудно-модулированной гармоник (когда период не постоянный, а медленно меняется) всегда будет растекание частоты.

2.3. Тренд: периодограмма тренда

Вообще, исходно периодограмма не предназначена для анализа тренда, так как тренд не периодический и при изменении длины ряда его разложение в сумму гармоник может сильно меняться.

Однако, при фиксированной длине ряда, вполне можно рассматривать приближение тренда суммой гармоник с низкими (маленькими) частотами.

Определение тренда как медленно-меняющейся компоненты (вторая производная небольшая) согласуется с этим. Действительно, во второй производной появляется коэффициент ω^2 , т.е. чем больше частота, тем больше вторая производная.

Итак, в периодограмме тренду соответствуют низкие частоты.

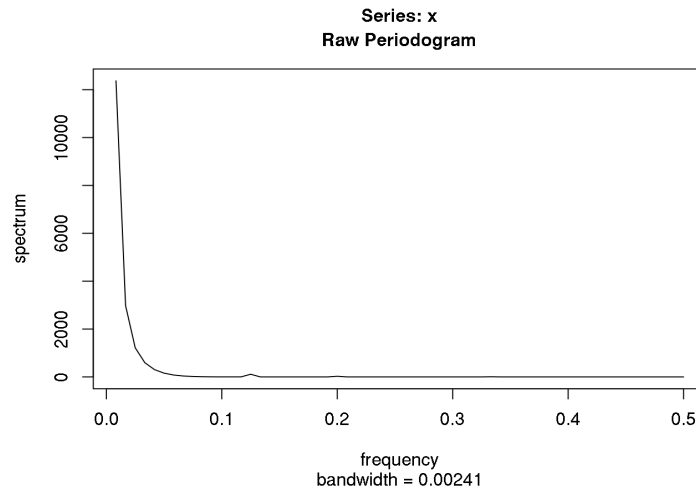


Рис. 2.4. Периодограмма тренда

А если в ряде есть и тренд, и периодика, то получим такую картинку: TODO.

Отсюда видно, что такие низкие частоты — как минимум, они должны быть меньше, чем частота периодической компоненты.

Замечание. Если тренд большой по сравнению с периодикой, то на графике пики от периодики не будут видны. Есть два способа борьбы с этим: либо рисование периодограммы в логарифмической шкале по оси Y, либо сначала вычитание тренда (имеется в виду линейный тренд) и потом рисование периодограммы остатка (обычно соотв. параметр называется DETREND).

2.4. Шум и его периодограмма

Шум — случайная составляющая временного ряда (например, стационарный процесс). Шум в данных может возникнуть из-за погрешности измерений или влияния случайных посторонних факторов.

$$x_n = s_n + \epsilon_n, \mathbb{E}\epsilon_n = 0.$$

Не будем давать строгое определение стационарного случайного процесса. Важно понимать, что стационарность означает, что свойства шума не меняются во времени.

Периодограмма стационарного случайного процесса — естественно, случайный набор чисел. Случ. процесс раскладывается по всем частотам.

Вид шума характеризуется математическим ожиданием периодограммы, т.е. тем, каков средний вклад тех или иных частот.

Если средний вклад всех частот одинаковый, то шум называется белым. Это справедливо для случайного процесса с нулевым математическим ожиданием, постоянной дисперсией и нулевой корреляцией между ξ_i и ξ_j , $i \neq j$. У красного шума чем меньше (ниже) частота, тем больше ее вклад.

Периодограмма белого шума выглядит следующим образом:

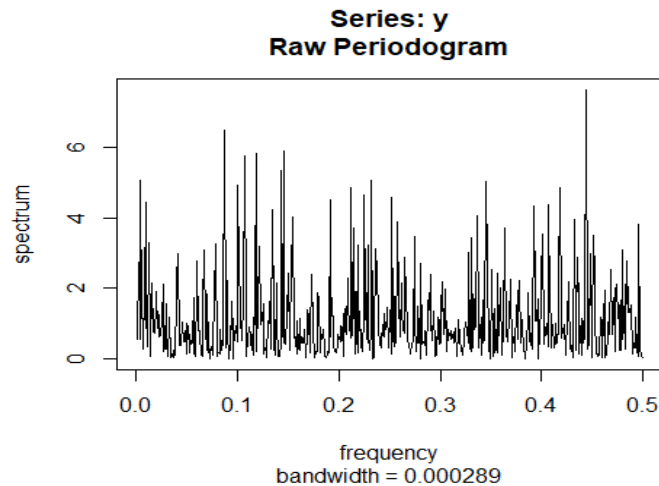


Рис. 2.5. Периодограмма белого шума

Свойство периодограммы белого шума: значение периодограммы имеет экспоненциальное распределение с одним и тем же средним. Почему важно, что экспоненциальное: у экспоненциального распределения длинный хвост, т.е. часто встречаются большие значения. Поэтому периодограмма такая 'лохматая'. Также это объясняет, почему иногда большой случайный пик от шума можно принять за гармонику.

Теперь можем пояснить, откуда такие названия. Ось частот можно интерпретировать и для цветовой гаммы (связано с длиной волны — больше длина волны, т.е. больше период и меньше частота). Каждому цвету соответствует свой набор частот с разной интенсивностью. Например, белый цвет состоит из смеси всех цветов (всех волн) одинаковой интенсивности. В красном цвете больше низких частот.

Другая интерпретация оси частот звуковая. Звуки также можно рассматривать как наборы звуковых волн разной частоты. Низкие звуки соответствуют низким частотам, высокие - высоким.

Пример периодограммы ряда с трендом, периодикой и шумом (TODO).

Заметим следующее. Средний квадрат значений белого шума (рассматриваем частный случай) стремится к его дисперсии, т.е. к константе. Одновременно, средний квад-

рат значений белого шума равен сумме значений периодограммы ()мы рассматриваем такую нормировку). Но при увеличении длины ряда число слагаемых (число узлов решетки с шагом $1/N$) увеличивается, поэтому средние значения периодограммы шума уменьшаются (и стремятся к нулю). Значение же пика для гармоники не меняется (без растекания). Поэтому чем длиннее ряд, тем проще в нем увидеть периодики.

Задание по лекции 2 может быть таким: есть графики рядов и есть графики периодограмм. Нужно поставить их в пары.

Глава 3

Лекция 3. Линейные фильтры

3.1. Линейный фильтр, импульсная характеристика, причинный фильтр, FIR

Очень часто встречающимся методом анализа временных рядов является сглаживание ряда с помощью скользящего среднего (усреднения значений ряда по скользящим отрезкам ряда). В виде формул это выглядит, например, так: $y_n = \sum_{i=-r}^r x_{n-i}/M$, где $M = 2r + 1$ (картинка). Видим, что результат является линейной комбинацией значений ряда. Это пример линейного фильтра с коэффициентами $1/M$. Оказывается, что есть теория, связанная с линейными фильтрами, которая позволяет предсказывать результат на основе коэффициентов фильтров, в частности, с точки зрения частотного содержания результата фильтрации, который рассматривался на предыдущей лекции (периодограммы).

Итак, введем определения.

Рассмотрим бесконечный временной ряд $\mathbf{X}^\infty = (\dots, x_{-1}, x_0, x_1, \dots)$. Тогда:

- $y_j = (\Phi(\mathbf{x}^N))_j = \sum_{i=-\infty}^{\infty} h_i \mathbf{x}_{j-i}^N$ — линейный фильтр;
- Набор коэффициентов $\{h_i\}$ — импульсная характеристика фильтра (impulse response)
- Если набор коэффициентов $\{h_i : h_i \neq 0\}$ конечен, то импульсная характеристика называется конечной (КИХ) — finite impulse response (FIR). Тогда $y_j = (\Phi(X))_j = \sum_{i=-r_1}^{r_2} h_i x_{j-i}$. Иначе фильтр называется БИХ (IIR). Примером фильтра БИХ является фильтр с экспоненциально убывающими коэффициентами. Мы будем рассматривать КИХ фильтры.
- Для прогноза очень важно, чтобы фильтр не заглядывал в будущее. Поэтому отдельно выделяют фильтры $y_j = (\Phi(\mathbf{x}^N))_j = \sum_{i=0}^{r_2} h_i \mathbf{x}_{j-i}^N$ — причинный фильтр (casual), смотрим только в прошлое. На этом примере видно, что, на самом деле, фильтр определяется не только коэффициентами, но и тем, с какой временной точкой соотносится результат фильтра.

На практике мы никогда не располагаем бесконечным рядом. Поэтому, при рассмотрении ряда конечной длины $\mathbf{X}_N = (x_1, \dots, x_N)$ можно сделать, например, следующее: вкладываем его в бесконечную последовательность $\mathbf{x}^\infty = (\dots, 0, \dots, 0, x_1, \dots, x_N, 0, \dots)$ и рассматриваем её как бесконечный временной ряд. // Вообще, много чего делается, это для примера, который подходит для временного ряда с нулевым средним.

3.2. Характеристики фильтра через его воздействие на $\cos(2\pi\omega n)$ (или на комплексную экспоненту) – АЧХ, ФЧХ

Напоминание: $y_j = (\Phi(\mathbf{x}^N))_j = \sum_{i=-\infty}^{\infty} h_i \mathbf{x}_{j-i}^N$ — линейный фильтр. Тогда:

- $H_\Phi(z) = \sum_i h_i z^{-i}$ — передаточная функция (transfer function);
- $|H_\Phi(e^{i2\pi\omega})| = A_\Phi(\omega)$ — амплитудно-частотная характеристика фильтра (АЧХ);
- $\phi_\Phi(\omega) = \text{Arg}(H_\Phi(e^{i2\pi\omega}))$ — фазово-частотная характеристика (ФЧХ).

Выглядит все очень страшно, однако нам достаточно всего одного простого факта:

Пусть исходный временной ряд имеет общий член $x_j = \cos(2\pi\omega j)$. Тогда:

$$y_j = (\Phi(\mathbf{x}^N))_j = A_\Phi(\omega) \cos(2\pi\omega j + \phi_\Phi(\omega)).$$

Так как линейный фильтр — линейное преобразование, то фильтр, примененный к сумме некоторых слагаемых (например, гармоник), равен сумме фильтров, примененных к каждому слагаемому по-отдельности.

Тут должно стать понятно, куда мы клоним: теперь мы знаем, как действует фильтр на разложение Фурье временного ряда.

Мы будем делать акцент на АЧХ, которая показывает, как меняется амплитуда в зависимости от частоты.

Вот представьте, что нам нужно сгладить ряд (сгладить = убрать быстрые колебания = оставить только колебания с низким частотами). Мы обсуждали, что гладкий тренд (или сигнал) соответствует низким частотам. Поэтому, чтобы сгладить ряд, нужен фильтр, у которого АЧХ на низких частотах большая, а на высоких — маленькая. У нас есть подозрение, что скользящее среднее должно быть как раз таким фильтром. Посмотрим на него с этой точки зрения.

3.3. АЧХ фильтра скользящего среднего, зависимость от длины окна

Скользящее среднее — частный случай линейного фильтра. Оно имеет один параметр M — ширина окна (порядок скользящего среднего) и коэффициенты (импульсную характеристику), равные $1/M$. Результат фильтра выглядит как среднее арифметическое $\sum_{i=1}^M x_{n-i}/M$, но с каким временем этот результата синхронизировать — это отдельный разговор. При нечетном M стандартный вариант — синхронизация с серединой. Про это поговорим позднее. Так как АЧХ зависит только от импульсной характеристики, то сейчас это не важно.

Характеристики фильтра:

- FIR: $h_i = 1/M, i = 0, \dots, M - 1$;
- $A_{\Phi}(\omega) = \frac{\sin(M\pi\omega)}{M \sin(\pi\omega)}, \omega \in [0, 0.5]$.

//Можно нарисовать при разных M АЧХ и результаты применения к шуму.

Видно, что чем больше M , тем быстрее АЧХ убывает (со скоростью $1/M$).

Отлично, мы это и хотели.

Но виден еще и дополнительный эффект — АЧХ равна нулю в точках k/M . Таким образом, скользящее среднее подавляет частоты периодик, которые имеют период, делящий длину окна нацело.

АЧХ скользящего среднего с длиной окна $M = 12$ (для усиления подавления шума можно взять и $M = 24, 36, 48, \dots$) для ряда, имеющего сезонность с периодом 12:

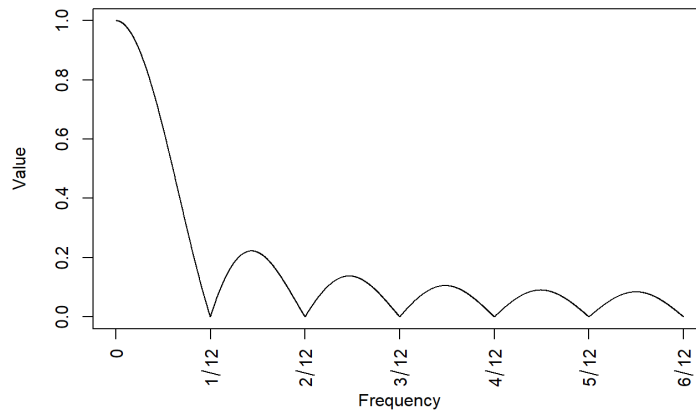


Рис. 3.1. АЧХ, $M = 12$

3.4. АЧХ фильтра перехода к разностям (дифференцирования)

Одной из стандартных операций, применяемых к временным рядам, является переход к последовательным разностям. Фильтр выглядит следующим образом:

$$y_n = x_n - x_{n-1},$$

т.е. импульсная характеристика равна $(1, -1)$.

Этот фильтр используется для устранения линейного тренда. Действительно, если был тренд $x_n = an + b$, то после применения фильтра получим $y_n \equiv a$. Однако, анализ АЧХ показывает, какую плату мы за это заплатим.

АЧХ такого фильтра:

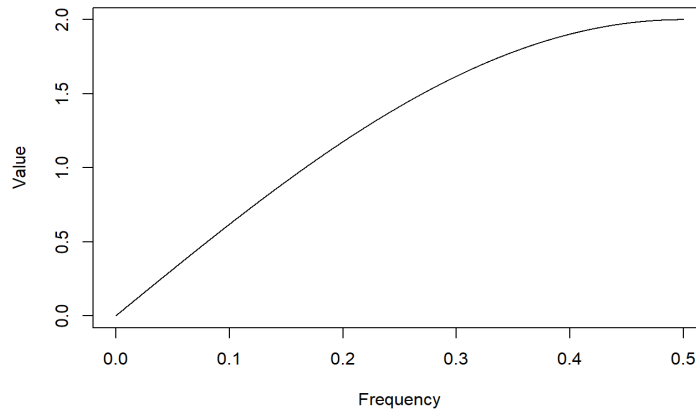


Рис. 3.2. АЧХ фильтра перехода к разностям

Действие этого фильтра обратное скользящему среднему: будут подавлены низкие частоты, полностью должна уйти константа, так как значение АЧХ в нулевой частоте равно 0 (что можно было предположить из вида фильтра — он напоминает формулу численного дифференцирования и поэтому переход к разностям часто так и называется дифференцированием). Таким образом, если ряд был суммой линейного тренда, периодической компоненты и шума, то тренд после дифференцирования действительно превращается в константу, но одновременно увеличивается шум (его высокочастотная часть) и даже возможно, что периодическая компонента перестанет быть видима на фоне шума.

//Замечание. Для броуновского движения дифференцирование справляется со ‘случайным трендом’ в том смысле, что броуновское движение (нестационарный процесс) превращается в белый шум (стационарный процесс).

3.5. Связь между периодограммами ряда до и после применения фильтра

//Может, перенести этот подраздел в начало, сразу после введения АЧХ?

Подытожим то, что ранее рассмотрели теоретически и на примерах.

Так как действие фильтра приводит к умножению амплитуды косинуса на значение АЧХ, периодограмма результата применения фильтра равна периодограмме исходного ряда, умноженной на квадрат АЧХ.

Рассмотрим периодограмму ряда без применения какого-либо фильтра (в т.ч. опция `detrend=FALSE`):

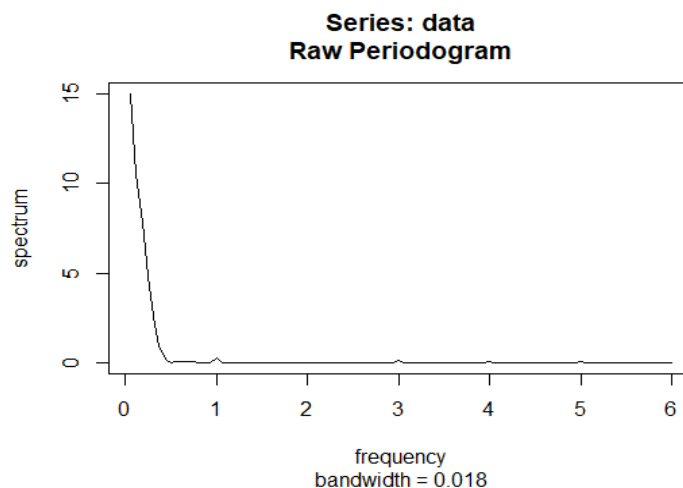


Рис. 3.3. Периодограмма ряда, `detrend = FALSE`

Видим большой вклад низких частот, то есть тренда, на фоне которого просто только проглядывается вклад периодики (сезонности). Применим к ряду фильтр скользящего среднего с $M = 12$, вычтем полученное из исходного ряда и для получившегося остатка построим периодограмму:

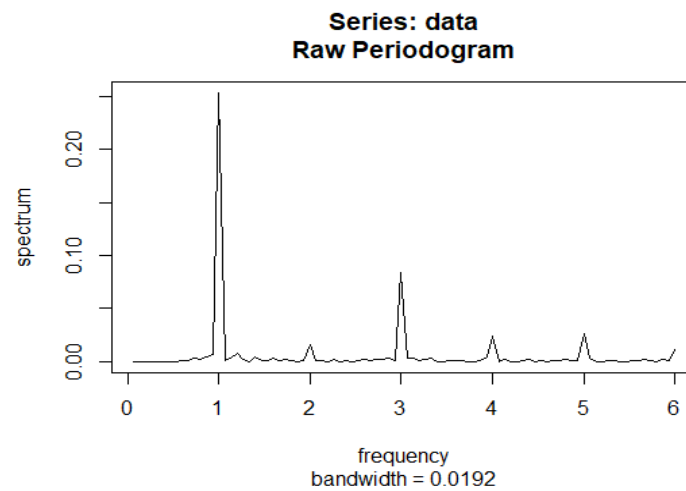


Рис. 3.4. Периодограмма остатка после вычитания результата применения фильтра, `detrend = FALSE`

Воздействие тренда полностью убрано, частоты периодик видны куда ярнее.

Построим периодограмму компоненты, непосредственно получившейся в результате сглаживания:

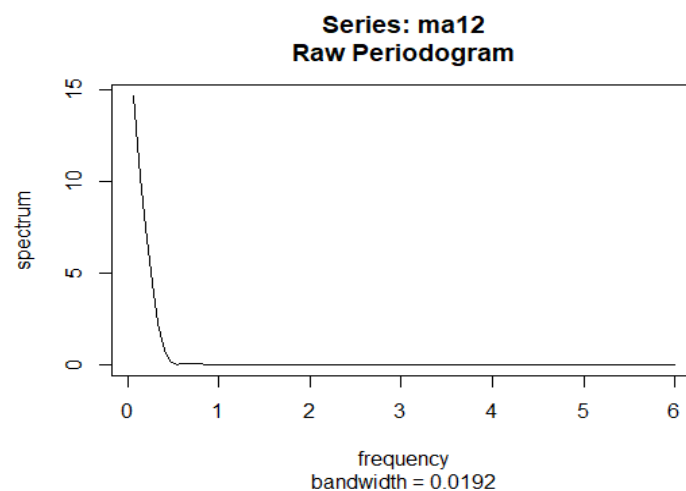


Рис. 3.5. Периодограмма фильтра, `detrend = FALSE`

Здесь наоборот, остались только низкие частоты (тренд), все остальные больше не видны.

P.S. Нужно заменить пример, чтобы тренд на периодограммах не забивал периодику.

3.6. Что такое запаздывание и отчего оно может возникать (на примере скользящего среднего)?

Мы уже упоминали, что результат применения фильтра можно соотносить с разными моментами во времени. Обсудим это более конкретно на примере скользящего среднего.

Пусть $M = 2k+1$ — нечетное. Вот есть у нас среднее арифметическое $\frac{1}{M} \sum_{i=-k}^k x_{n-i}$.

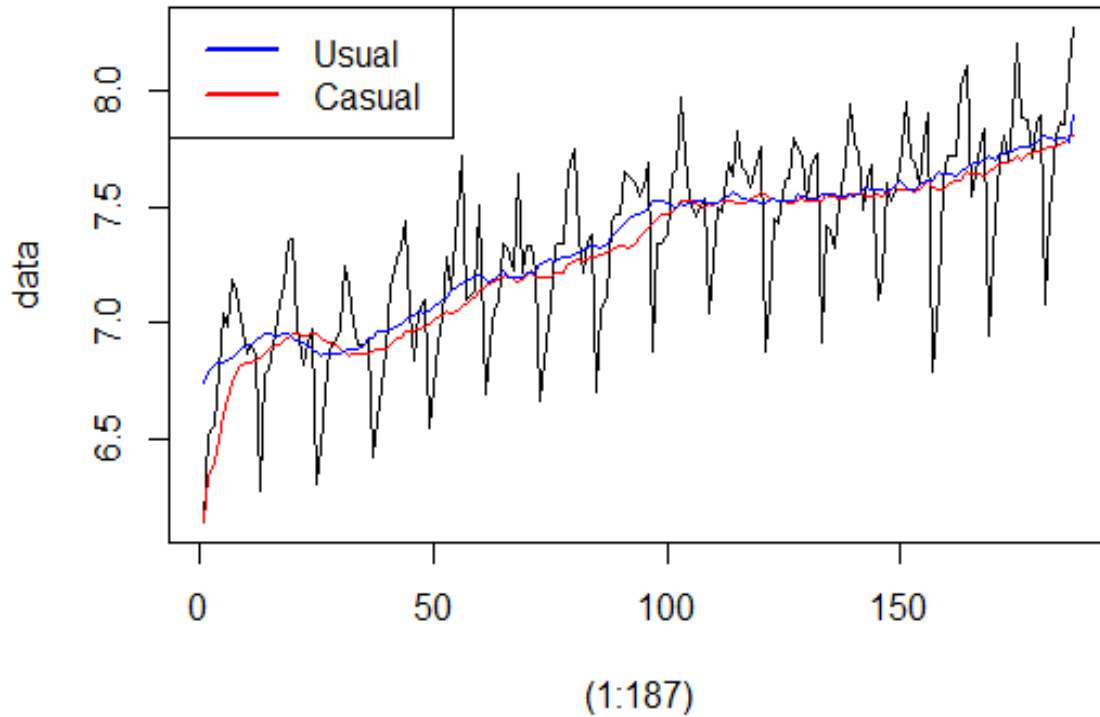
Есть два вполне осмысленных варианта: $y_n^{mid} = \frac{1}{M} \sum_{i=-k}^k x_{n-i}$ и $y_{n+k}^{last} = \frac{1}{M} \sum_{i=-k}^k x_{n-i}$.

В первом случае легко увидеть, что линейная функция остается без изменения: $y_n^{mid} = x_n$ при $x_n = an + b$, $n = k+1, \dots, n-k$ (говорят, что нет смещения во времени). Обратите внимание, что последние точки не описываются этим фильтром. Также (на самом деле, это фактически то же самое утверждение), этот фильтр не является причинным (causal), так как заглядывает в будущее на k точек.

//Наверно, тут можно выписать ФЧХ, но я ей не разбиралась, как она выглядит.

Во втором случае фильтр получается причинным (causal), однако очевидно, что он запаздывает: $y_n^{last} = y_{n-k}^{mid} = a(n-k) + b$.

Сравним y_n^{mid} и y_n^{last} на примере.

Рис. 3.6. МА, $M = 12$, $T = 12$

Видно, что причинный фильтр повторяет обычное скользящее среднее с запаздыванием (запаздывание равно $M/2$). (//ничего тут не видно, нужен пример лучше)

Если $M = 2k$ четное, то сделать скользящее среднее несмещенным по времени невозможно. С другой стороны, четное M очень нужно, например, чтобы убрать сезонность с периодом 12 (в данных по месяцам).

Оказывается, можно слегка изменить коэффициенты фильтра, не испортив его свойств.

Определим

$$z_n = \frac{1}{M} \sum_{i=-k}^{k-1} x_{n-i}.$$

Теперь применим фильтр порядка 2 к полученному ряду:

$$y_n = \frac{z_{n-1} + z_n}{2}.$$

Перепишем все через значения исходного ряда:

$$y_n = \frac{x_{n-k} + x_{n+k}}{2M} + \frac{1}{M} \sum_{i=-k}^{k-1} x_{n-i}.$$

Теперь окно стало симметричным относительно n . Например, для устранения сезонности получим фильтр с коэффициентами $1/24$, потом 11 по $1/12$ и в конце снова $1/24$.

Далее идет дополнительный материал.

3.7. Фильтр как оценка сигнала

Посмотрим на фильтры с другой стороны (не с частотной). С одной стороны, сглаживание можно рассматривать как удаление высокочастотных составляющих. С другой стороны, если мы наблюдаем зашумленный тренд/сигнал, то хочется удалить шум и оставить только сигнал. Задачи очень похожие, но разница есть.

Рассмотрим последовательность следующего вида: $x_n = s_n + \epsilon_n$, причем $\mathbb{E}\epsilon_n = 0$, $\mathbb{D}\epsilon_n = \sigma^2$, ϵ_i — независимы. Рассмотрим следующий линейный фильтр:

$$y_n = \sum_{i=-r_1}^{r_2} h_i x_{n-i}.$$

Итак, посмотрим на фильтры с точки зрения того, как результат фильтрации оценивает сигнал. Как обычно, точность оценки измеряется с помощью $\text{MSE} = \text{дисперсия} + (\text{смещение в квадрате})$.

Рассмотрим слагаемые по-отдельности.

3.7.1. Дисперсия

Итак, рассмотрим оценку сигнала $\hat{s}_n = \sum_{i=-r_1}^{r_2} h_i x_{n-i}$. Посчитаем дисперсию этой оценки:

$$\mathbb{D}\hat{s}_n = \mathbb{D} \sum_{i=-r_1}^{r_2} h_i (s_i + \epsilon_i) = \mathbb{D} \sum_i h_i \epsilon_i = \sum_i h_i^2 \sigma^2 = \sigma^2 \sum_i h_i^2.$$

Итак, чтобы минимизировать дисперсию, нужно требовать $\sum_i h_i^2 = \|\mathbf{h}\|^2 \rightarrow \min$.

Для скользящего среднего с шириной окна M $\mathbb{D}\hat{s}_n = \frac{\sigma^2}{M}$, т.е. чем больше окно, тем меньше дисперсия у результата.

Если предположить, что для сигнала выполняется следующее: $s_n = \sum_{i=-r_1}^{r_2} h_i s_i$, т.е. фильтр, примененный к сигналу, его не меняет, то смещение равно нулю ($\mathbb{E}\hat{s}_n = s_n$) и

результат полностью определяется дисперсией. (Но даже в этом случае брать большую ширину окна может быть плохо для конечного ряда, так как слишком много крайних точек фильтром не будет описываться.) Вообще, несмещенность может быть довольно ограничительным предположением. Мы видели, что скользящее среднее обладает таким свойством (только?) для линейного сигнала.

3.7.2. Смещение, роль второй производной

//Этот раздел сложен тем, что приходится переходить к непрерывному времени. Поэтому я и думаю, что этот раздел не включать, хотя концептуально раздел важный. Но для 3 семестра можно обойтись без концепций. Во всяком случае, пришедшие магистры это не смогли понять.

Рассмотрим непрерывный аналог фильтрации, примененный к сигналу:

$$y(a) = \int_{-\delta}^{\delta} s(a+x)w(x)dx,$$

где $w(x)$ — весовая функция, причем $\int_{-\delta}^{\delta} w(x)dx = 1$ (веса нормированы) и $\int_{-\delta}^{\delta} xw(x)dx = 0$ (например, функция $w(x)$ симметричная). Здесь δ вместо ширины окна M , а весовая функция w вместо коэффициентов h_i .

Под интегралом разложим функцию s в ряд Тейлора (предполагаем, что s — гладкая):

$$y(a) \approx s(a) \times 1 + s'(a) \times 0 + \frac{s''(a)}{2} \int_{-\delta}^{\delta} x^2 w(x)dx.$$

Таким образом, смещение $y(a) - s(a) \approx \frac{s''(a)}{2} \int_{-\delta}^{\delta} x^2 w(x)dx$.

Если скользящее среднее с равными весами, то $w(x) = \frac{1}{2\delta}$ и

$$y(a) \approx s(a) + \frac{s''(a)}{4\delta} \int_{-\delta}^{\delta} x^2 dx = s(a) + \frac{s''(a)\delta^2}{6},$$

то есть смещение зависит от второй производной. Знак смещения зависит, соответственно, от знака второй производной, то есть от характера перегиба функции s . Понятно, что такой фильтр не имеет смысла, если вторая производная большая (и δ недостаточно маленькое).

Пример. <Картинка с косинусом и небольшой шириной окна, где видно, что на максимумах (вторая производная отрицательная) результат сглаживания меньше, а на минимумах больше.>

Таким образом, как обычно имеет trade-off между смещением и дисперсией при выборе ширины окна.