

Санкт-Петербургский государственный университет  
Прикладная математика и информатика  
Статистическое моделирование

Набрано: Федоров Никита, Познизова Вероника

## СТАТИСТИЧЕСКИЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Конспект лекций/практики Голяндиной Н.Э.

Санкт-Петербург

2021

# Оглавление

|                                                                                                                                |          |
|--------------------------------------------------------------------------------------------------------------------------------|----------|
| <b>Глава 1. Общая часть . . . . .</b>                                                                                          | <b>3</b> |
| Введение . . . . .                                                                                                             | 5        |
| 1.1. Периодограмма временного ряда . . . . .                                                                                   | 7        |
| 1.2. Тренд: периодограмма тренда . . . . .                                                                                     | 8        |
| 1.3. Периодическая компонента и ее периодограмма . . . . .                                                                     | 9        |
| 1.4. Растекание частоты в периодограмме и методы его устранения . . . . .                                                      | 10       |
| 1.5. Шум и его периодограмма . . . . .                                                                                         | 11       |
| 1.6. Периодограмма как оценка спектральной плотности, распределение значений, сглаживание периодограммы . . . . .              | 12       |
| 1.7. Различия между сглаживанием и выделением тренда . . . . .                                                                 | 13       |
| 1.8. Линейный фильтр, импульсная характеристика, причинный фильтр, FIR . . . . .                                               | 14       |
| 1.9. Характеристики фильтра через его воздействие на $\cos(2\pi\omega n)$ (или на комплексную экспоненту) – АЧХ, ФЧХ . . . . . | 15       |
| 1.10. АЧХ фильтра скользящего среднего, зависимость от длины окна . . . . .                                                    | 16       |
| 1.11. АЧХ фильтра перехода к разностям (дифференцирования) . . . . .                                                           | 18       |
| 1.12. Что такое запаздывание и отчего оно может возникать (на примере скользящего среднего)? . . . . .                         | 19       |

## Глава 1

### Общая часть

- Что такое периодограмма?
- Что такое тренд (разные варианты определения)? Как выглядит периодограмма тренда?
- Что такое периодическая компонента? Как выглядит ее периодограмма?
- Растекание частоты в периодограмме. Подправка длины ряда для ее устранения.
- Что такое шум? Как выглядит его периодограмма?
- Периодограмма как оценка спектральной плотности. Распределение значений. Сглаживание периодограммы.
- Чем отличается сглаживание от выделения тренда?
- Что такое линейный фильтр, импульсная характеристика? Причинный фильтр, FIR.
- Характеристики фильтра через его воздействие на  $\cos(2\pi\omega n)$  (или на комплексную экспоненту) – АЧХ, ФЧХ.
- АЧХ фильтра скользящего среднего, зависимость от длины окна.
- АЧХ фильтра перехода к разностям (дифференцирования).
- Что такое запаздывание и отчего оно может возникать (на примере скользящего среднего)?
- Смещение при сглаживании фильтром скользящего среднего. Роль второй производной.
- Фильтр для подавления шума. Роль нормы коэффициентов фильтра.
- Как связаны периодограммы ряда до применения фильтра и после применения фильтра?

- Модели данных – аддитивная и мультипликативная.
- Методы стабилизации дисперсии в разных моделях (логарифмирование, извлечение квадратного корня, ...).
- Выделение тренда у ряда с сезонностью (выбор длины окна в скользящем среднем).
- Переход к разностям – плюсы и минусы (устранение тренда, превращение ряда в стационарный, усиление вклада высоких частот).
- Скользящее среднее и скользящая медиана.
- Выделение тренда с помощью параметрической регрессии.
- Выделение тренда с помощью метода LOESS
- Нахождение огибающей периодического ряда с помощью выделения тренда.
- Оценивание поведения дисперсии шума с помощью выделения тренда.
- Метод разложения Classical seasonal decomposition.
- Метод разложения STL.

## 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).

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

**Первый тип рядов: тренд + периодическая компонента + шум = сигнал + шум.**  $x_n = t_n + p_n + \epsilon_n = s_n + \epsilon_n, \mathbb{E}\epsilon = 0$ .

Сигнал — неслучайная (детерминированная) составляющая ряда, в нем и содержится структура.

Тренд: определения, примеры, зависимость от таймфрейма.

Периодическая компонента с периодом  $T$ : то, что повторяется каждые  $T$  точек.

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.

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

$A$  — амплитуда.

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

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

Сумма косинусов с периодами 12, 12/2, 12/3, 12/4, 12/5, 12/6.

Относительность понятия тренда.

Если нет периодичности, то сигнал = тренд. Сигнал можно определить как мат.ож. временного ряда.

Тренд может быть детерминированным, а может быть случайным

**Второй тип рядов: ряд — случайный процесс** Случайный процесс (с дискретным временем)  $\xi_1, \dots, \xi_N$  состоит из последовательности случайных величин. Т.е. это многомерный случайный вектор. Самый стандартный класс случайных процессов — стационарные случайные процессы. Стац. процессы в слабом смысле:  $\xi_k$  имеют одинаковое мат.ож. (часто равно нулю), одинаковую дисперсию и корреляция/ковариация зависит только от разницы во времени:  $\text{Cov}(\xi_t, \xi_s) = V(t - s)$ .

Например,  $x_n = \epsilon_n$ ,  $\mathbb{E}\epsilon_n = 0$ ,  $\epsilon_n$  одинаково распределены и независимы между собой (iid — independent, identically distributed). Это белый шум. Структура? Предсказание?

Другой пример:  $x_n = \phi x_{n-1} + \sigma \epsilon_n$ . Вопрос: если  $\sigma = 0$ , то как ведет себя ряд при разных  $\phi$ ?

Оказывается, что если  $|\phi| < 1$  и добавить белый шум на каждом шаге, то получим стационарный случайный процесс. Если  $\phi$  еще и больше 0, то этот процесс называется красным шумом. За счет чего такие процессы можно прогнозировать?

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  $\sqrt{n}$ .

## 1.1. 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  $N$ .

Нам интересно разложить функцию ряда по этому базису, чтобы определить влияние конкретных частот на значения ряда. Пусть

$$f : \{1, \dots, N\} \rightarrow \mathbb{R}$$

функция, которая задаёт наш ряд. Нам известны значения этой функции, и мы хотим получить представление:

$$f(n) = c_0 + \sum_{k=1}^{\lfloor \frac{N}{2} \rfloor} \{c_k \cdot \cos\left(2\pi n \frac{k}{N}\right) + s_k \cdot \sin\left(2\pi n \frac{k}{N}\right)\},$$

где  $n \in \{1, \dots, N\}$ . Нахождение коэффициентов  $\{c_k\}, \{s_k\}$  можно рассмотреть как задачу регрессии. Эквивалентное представление разложения Фурье:

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

где  $A_k^2 = c_k^2 + s_k^2$ ,  $A_0 = c_0$ .

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

$$\Pi\left(\frac{k}{N}\right) = \frac{N}{2} \begin{cases} c_0^2 & \text{if } k = 0 \\ c_{\frac{N}{2}}^2 & \text{if } N \text{ is even and } k = \frac{N}{2} \\ c_k^2 + s_k^2 & \text{otherwise} \end{cases}$$

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

В R периодограмму можно построить с помощью функции `spec.pgram`.

## 1.2. Тренд: периодограмма тренда

Если воспринимать тренд как медленно-меняющуюся компоненту (вторая производная небольшая), то можно это переформулировать в терминах частот. Тренду в разложении ряда соответствуют низкие (маленькие) частоты.

(Посчитайте, как зависит вторая производная от частоты.)

В периодограмме тренду соответствуют низкие частоты.

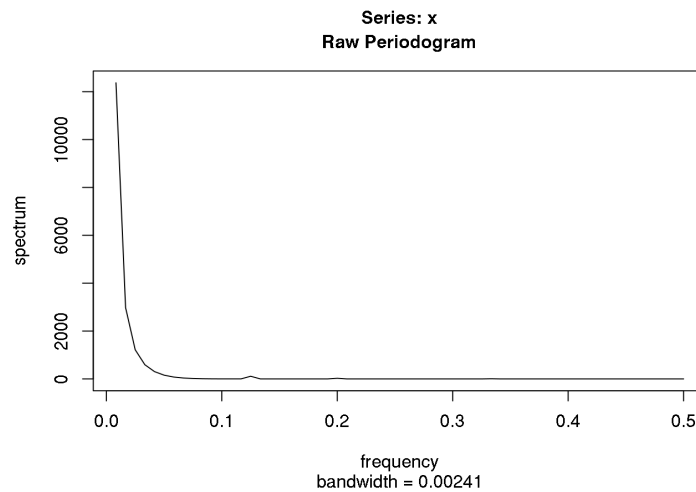


Рис. 1.1. Периодограмма ряда без вычитания тренда (опция `detrend=F`)

Увидеть частоты гармоник на графике довольно сложно.

### 1.3. Периодическая компонента и ее периодограмма

Периодическая компонента — сумма гармоник ряда, периодически повторяющаяся его компонента. Чем больше амплитуда у гармоники, тем больше значение мощности частоты на периодограмме.

Мощности низких частот (тренд) обычно подавляют на графике частоты гармоник. Чтобы исправить ситуацию, при первичном анализе ряда в `spec.pgram` можно использовать параметр `detrend`, чтобы избавиться от тренда в первом приближении (вычитается линейный тренд).

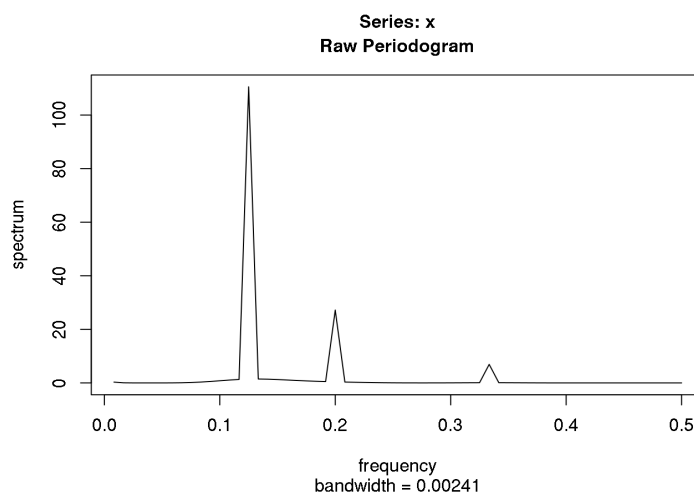


Рис. 1.2. Периодограмма ряда с вычитанием тренда (опция `detrend=T`)

## 1.4. Растекание частоты в периодограмме и методы его устранения

Если длина ряда не будет кратна частоте периодики, то мы не сможем её выделить на графике, и получим растекание частоты (frequency leakage).

Для примера рассмотрим ряд длиной 105 точек:  $y_i = \cos(\frac{2\pi x_i}{10})$ . Построим его периодограмму:

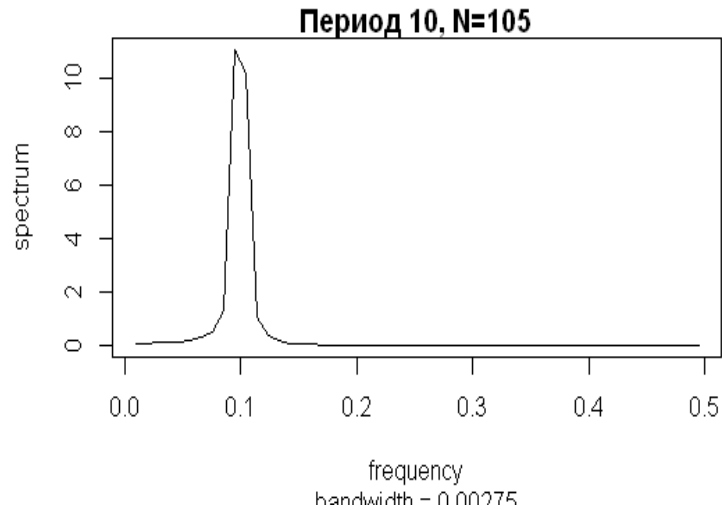


Рис. 1.3. Периодограмма исходного ряда

Первый вариант: ряд с тем же периодом и длиной 100 (нужно использовать `taper = 0` — спец.параметр, который что-то улучшает, но также искажает периодограмму; если поставить 0, то будет чистый ноль везде кроме пиков):

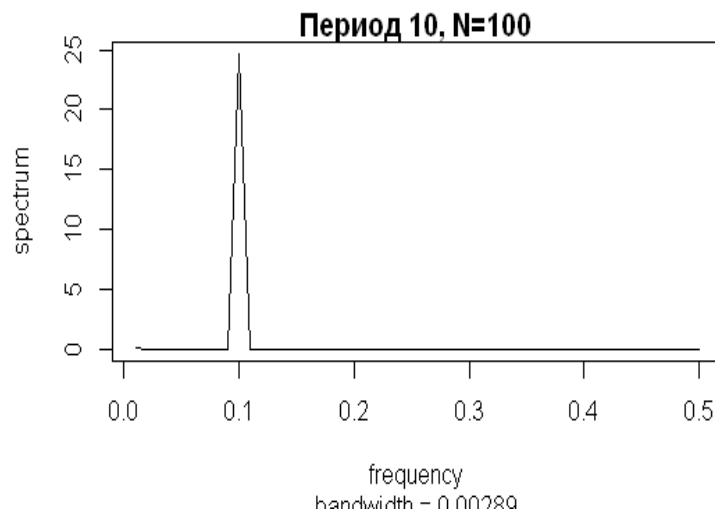


Рис. 1.4. Коррекция периодограммы

## 1.5. Шум и его периодограмма

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

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

Периодограмма белого шума (случайный процесс с нулевым математическим ожиданием, постоянной дисперсией и некоррелированный во времени) выглядит следующим образом:

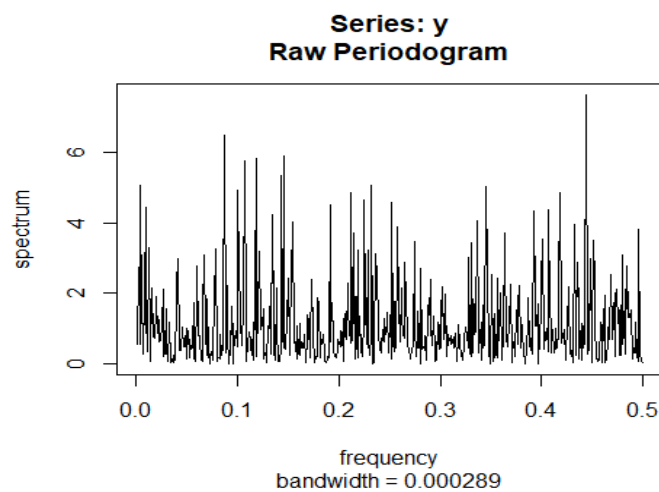


Рис. 1.5. Периодограмма белого шума

Однако, стационарный процесс характеризует спектральная плотность, которая равна математическому ожиданию периодограммы (с соотв. нормировкой). Спектральная характеристика случ.процесса показывает, какие частоты входят в процесс с БОЛЬШИМ вкладом, какие с меньшим.

Отсюда названия процессов: белый шум — ... красный шум - ...

Периодограмма — несмещенная оценка случ.шума. Однако, она не является состоятельной оценкой, ее дисперсия не стремится к нулю.

(Увеличиваем  $N$ , решетка периодограммы становится чаще, но значения все так же скачут вокруг значений спектральной плотности.)

Поэтому для оценки спектральной плотности периодограмму сглаживают взвешенным скользящим средним, о котором расскажу немного позднее.

## **1.6. Периодограмма как оценка спектральной плотности, распределение значений, сглаживание периодограммы**

Если процесс стационарный, периодограмма является оценкой спектральной плотности ряда (впрочем, эта оценка не является состоятельной). Для анализа рядов вроде красного/белого шума имеет смысл рассматривать сглаженную периодограмму. Иначе, если в ряде есть сезонность, то при сглаживании можно случайно убрать пики. В R в функции `spec.pgram` сглаживание осуществляется с помощью скользящего среднего.

## 1.7. Различия между сглаживанием и выделением тренда

В результате сглаживания действительно получится медленно меняющаяся компонента. Если понимать тренд как некоторую неслучайную функцию, то можем получить тренд. Если под трендом подразумевается среднее, то тренд не получим.

## 1.8. Линейный фильтр, импульсная характеристика, причинный фильтр, FIR

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)
- набор коэффициентов  $\{h_i : h_i \neq 0\}$ , то есть ненулевые коэффициенты импульсной характеристики — finite impulse response (FIR). Тогда  $y_j = (\Phi(X))_j = \sum_{i=-r_1}^{r_2} h_i x_{j-i}$
- фильтр  $y_j = (\Phi(\mathbf{x}^N))_j = \sum_{i=0}^{r_2} h_i \mathbf{x}_{j-i}^N$  — причинный фильтр (casual), смотрим только в прошлое.

На практике мы никогда не располагаем бесконечным рядом. Поэтому, при рассмотрении ряда конечной длины  $\mathbf{X}_N = (x_1, \dots, x_N)$  делается 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.

### 1.9. 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).

Пусть  $\mathbf{x}^N = \cos(2\pi\omega j)$ . Тогда:

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

АЧХ показывает как меняется амплитуда в зависимости от частоты.

### 1.10. АЧХ фильтра скользящего среднего, зависимость от длины окна

Скользящее среднее — частный случай линейного фильтра. Оно имеет один параметр  $M$  — ширина окна (порядок скользящего среднего). Для  $M = 2k + 1$  определим результат применения фильтра к ряду  $X_N$  следующим образом:

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

Таким образом, если ширина окна является нечетным числом, проблем не возникает. Если длина окна является четным числом, то процедура применения фильтра немного усложняется:

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

Теперь применим фильтр порядка 2 к полученному ряду:

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

Перепишем все через значения исходного ряда:

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

Теперь окно стало симметричным.

Характеристики фильтра:

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

Таким образом, скользящее среднее подавляет частоты периодик, которые имеют период, делящий длину окна нацело. Вообще, чем больше окно, тем сильнее скользящее среднее подавляет высокие частоты.

АЧХ скользящего среднего с длиной окна  $M = 12$  для ряда, имеющего сезонность с периодом 12:

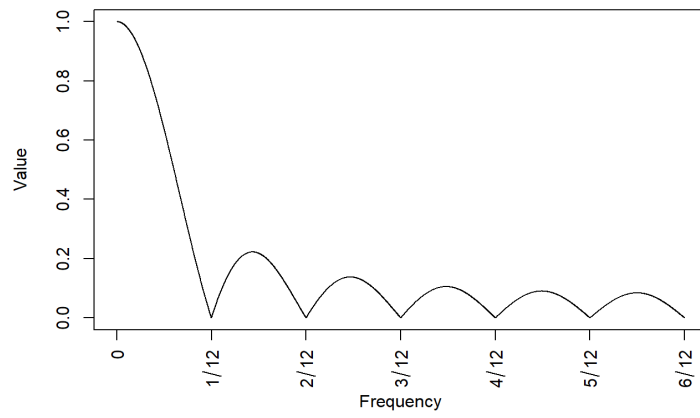


Рис. 1.6. АЧХ,  $M = 12$

### 1.11. АЧХ фильтра перехода к разностям (дифференцирования)

Фильтр выглядит следующим образом:

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

АЧХ такого фильтра:

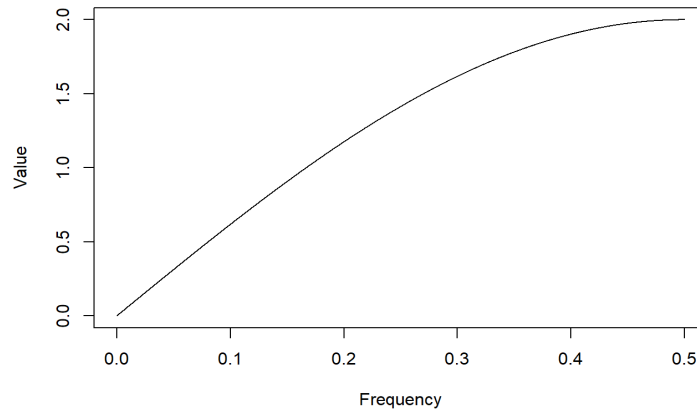


Рис. 1.7. АЧХ фильтра перехода к разностям

Действие этого фильтра обратное скользящему среднему: будут подавлены низкие частоты, полностью должна уйти константа (что можно было предположить из вида фильтра — он напоминает формулу численного дифференцирования).

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

Рассмотрим какой-нибудь причинный фильтр, например скользящее среднее, которое считается только по «прошлому». Сравним его с обычным скользящим средним:

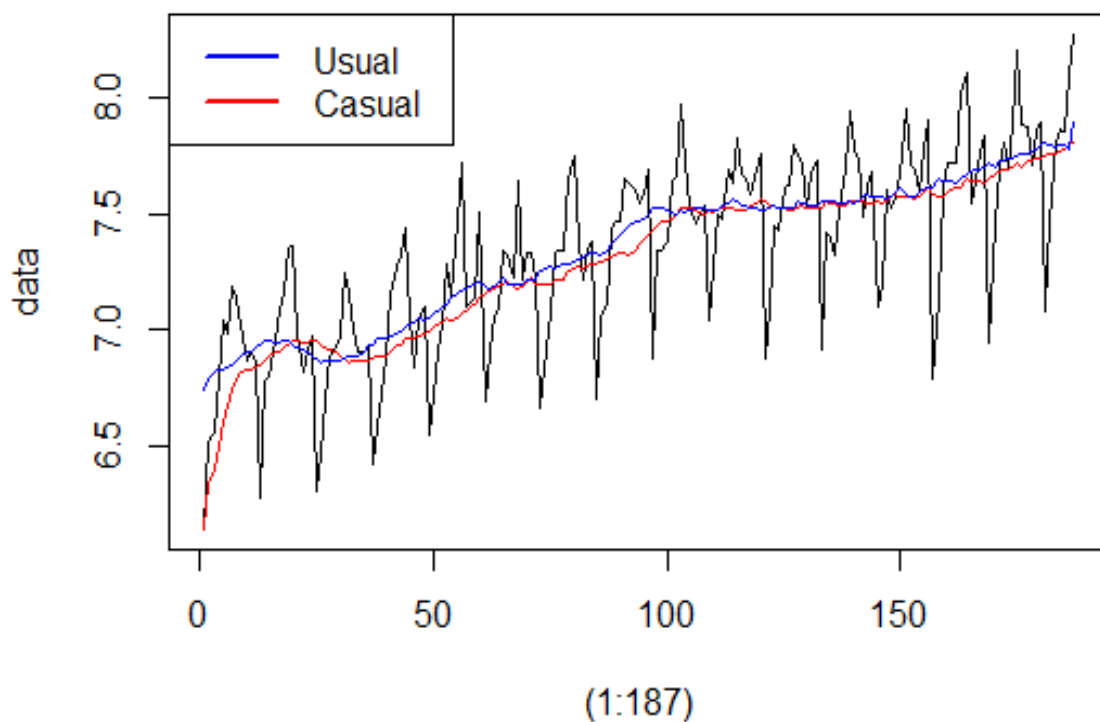


Рис. 1.8. МА,  $M = 12$ ,  $T = 12$

Видно, что причинный фильтр повторяет обычное скользящее среднее с запаздыванием (запаздывание равно  $M/2$ ).