

Пропуски, аномалии и структурные сдвиги

Заполнение пропусков

Заполнение пропусков: план

- Линейная интерполяция.
- Модели для заполнения пропусков.
- Использование STL-разложения.

Линейная интерполяция

Идея

Заполним пропуски так, чтобы восстановленные значения идеально ложились на прямую (образовывали арифметическую прогрессию),

$$\Delta y_t^{imp} = const.$$

Пример:

10, NA, NA, 100.

10, 40, 70, 100

Модели для заполнения пропусков

1. Оцениваем модель, **допускающую** пропуски в данных. ARIMA подходит! И автоматическая ARIMA тоже!
2. Пропущенные значения y_t заменяем на условное математическое ожидание, полагая оценённые параметры модели равным истинными,

$$y_t^{imp} = \mathbb{E}(y_t \mid \text{данные}).$$

Используется **фильтр Калмана**.

Возможность оценивать модель на данных с пропусками сильно зависит от **реализации**.

Использование STL-разложения

1. Раскладываем ряд с пропусками на составляющие:

$$y_t = \text{trend}_t + \text{seasonal}_t + \text{remainder}_t = \text{seasonal}_t + \text{deseason}_t.$$

STL восстанавливает **сезонную компоненту** без пропусков!

2. Восстанавливаем пропущенные значения десезонированного ряда **линейной** интерполяцией.
3. Пропущенные значения y_t заменяем на сумму восстановленных десезонированных значений и сезонной составляющей,

$$y_t^{imp} = \text{seasonal}_t + \text{deseason}_t^{imp}.$$

Зачем заполнять пропуски?

- Иногда заполнение пропусков — основная задача.
- Возможность использовать больше алгоритмов прогнозирования для восстановленного ряда.
- Возможность использовать восстановленный ряд как предиктор.

Заполнение пропусков: итоги

- Линейная **интерполяция**: просто и быстро!
- Использование **ARIMA** или более сложных моделей.
- **STL-разложение** и восстановление компонент.
- **Вариации** у каждого алгоритма.

Обнаружение аномалий

Обнаружение аномалий: план

- Какое наблюдение считать аномальным?
- Алгоритмы обнаружения и исправления аномалий.
- Зачем искать аномальные наблюдения?

Какое наблюдение считать аномальным?

Деление наблюдений на аномальные и обычные
субъективно.

Неформально, аномальное наблюдение выбивается из
основной динамики ряда.

Что считать «основной динамикой»? Что значит
«выбивается»?

Алгоритм обнаружения аномалий

- Берём любой алгоритм, позволяющий выделять из ряда **остаток** \hat{u}_t .

Подойдут как модели *ARIMA*, *ETS*, ..., так и алгоритм *STL*.

Остатком для моделей называют разницу фактическим значением и прогнозом внутри обучающей выборки.

- Оцениваем **стандартную ошибку** остатков.
- Если остаток по модулю больше **трех** стандартных ошибок, считаем наблюдение аномальным.

Исправление аномалий

Вычитаем из аномального наблюдения остаток:

$$y_t^{imp} = y_t - \hat{u}_t.$$

Зачем искать аномальные наблюдения?

- Иногда обнаружение аномалий — **основная задача**.
- Возможность получить **более точные** прогнозы для исправленного ряда.
- Возможность получить **более точные** прогнозы, если использовать исправленный ряд как предиктор.

Обнаружение аномалий: итоги

- Берём любой алгоритм (STL, ARIMA, ETS, ...), выделяющий из ряда **остаток**.
- Есть **куча** специальных алгоритмов.
- Если остаток **велик**, то считаем наблюдение аномальным.
- Чтобы исправить аномальное наблюдение, заменяем остаток **на ноль**.
- **Исправление** аномальных наблюдений перед прогнозированием может улучшить прогнозы!

Обнаружение структурного сдвига

Обнаружение структурного сдвига: план

- Что такое структурный сдвиг?
- Обнаружение одного структурного сдвига.
- Обнаружение нескольких структурных сдвигов.

Что считать структурным сдвигом?

Деление временного ряда на периоды между структурными сдвигами **субъективно**.

Неформально, момент структурного сдвига **меняет** поведение ряда.

Что считать «меняет»?

Идея обнаружения отдельного сдвига

- Стартуем со штрафной функции, измеряющей **неоднородность** наблюдений y_a, y_{a+1}, \dots, y_b ,

$$C(y_{a:b}).$$

- Перебираем все моменты $\tau \in [1; T - 1]$ и находим минимум величины

$$C(y_{1:\tau}) + C(y_{\tau+1:T}).$$

Подозреваем, что сдвиг мог быть в этот момент τ^* .

- Считаем, что сдвиг был в τ^* , если суммарная неоднородность фрагментов **сильно** меньше неоднородности всего ряда,

$$C(y_{1:\tau^*}) + C(y_{\tau^*+1:T}) < C(y_{1:T}) - \beta.$$

Выбор штрафной функции C

- Есть **огромное** количество вариантов.
- Часто берут функцию лог-функцию правдоподобия **некоторой** модели, домноженную на минус два:

$$C(y_{a:b}) = -2 \max_{\theta} \ln L(y_a, \dots, y_b \mid \theta).$$

Простейший вариант: считать, что $y_t \sim \mathcal{N}(\mu, \sigma^2)$ и независимы.

- Выбор функции C связан с выбором β при проверке наличия сдвига в подозрительной точке τ^* ,

$$C(y_{1:\tau^*}) + C(y_{\tau^*+1:T}) < C(y_{1:T}) - \beta.$$

Чем больше параметров в θ , тем больше должно быть β .

Как обнаружить много структурных сдвигов?

- Запустить алгоритм по обнаружению **одного** структурного сдвига.
Если алгоритм не обнаружил сдвиг, то считаем, что сдвигов на данном участке нет.
- Разбиваем исходный ряд на **два участка** согласно обнаруженному структурному сдвигу.
- **Рекурсивно** запускаем алгоритм обнаружения одного структурного сдвига на **каждом** обнаруженном участке.

Преобразования для поиска сдвига

Сдвиг может **легче** обнаруживаться на преобразованном ряду.

- **Простые действия** с исходным рядом: логарифм, преобразование Бокса-Кокса, переход к разностям.
- Разложение на компоненты и **поиск сдвига в компонентах** ряда: *STL*, *ETS*, ...

Зачем искать структурные сдвиги?

- Иногда обнаружение сдвигов — **основная задача**.
- Возможность получить **более точные** прогнозы, если добавить в предикторы лаг-переменную равную единице после сдвига.
- Возможность получить **более точные** прогнозы других рядов, если скорректировать структурный сдвиг в предикторе.

Обнаружение структурного сдвига: итоги

- Есть **куча** специальных алгоритмов.
- Чтобы найти **много** сдвигов, достаточно поискать очередной сдвиг на уже выявленных участках ряда.
- *STL* разложение позволяет искать **сдвиги в компонентах** ряда.

Байесовские структурные модели

Байесовские структурные модели: план

- Как добавить предикторы в ETS ?
- Идея байесовского подхода.
- Пример структурной модели.

Как добавить предикторы в *ETS*?

- Классическая *ETS* модель **не позволяет** включать предикторы.
- А **что мешает** их туда добавить и получить новую модель?
В новой модели может оказаться **слишком много** параметров.
Качество прогнозов может быть плохим.
- Спасительная идея — **регуляризация**.
Рассматриваем модель с большим числом параметров и дополнительной информацией, что параметры **небольшие**.

Байесовский подход!

- Трактуем все параметры как ненаблюдаемые **случайные величины**, $\theta = (a, b, c)$.
- **Модель** задаёт распределение ряда при заданных параметрах,

$$y_t = a + u_t + bu_{t-1}, \quad u_t \sim \mathcal{N}(0; c).$$

- Добавляем информацию в виде **априорного распределения**,

$$a \sim \mathcal{N}(0; 100), \quad b \sim \mathcal{N}(0, 1), \quad \ln c \sim \mathcal{N}(0; 4).$$

- Алгоритмы MCMC (Markov Chain Monte Carlo) позволяют сгенерировать большую выборку из **апостериорное распределение**

$$(a, b, c \mid y_1, y_2, \dots, y_T).$$

Немного про МСМС

- Выборка из апостериорного распределения $(\theta \mid y)$ позволяет считать **всё!**

$$(a_1, b_1, c_1), (a_2, b_2, c_2), \dots, (a_S, b_S, c_S).$$

- Имея значения параметров можно симулировать **будущие траектории**.
- МСМС позволяет работать с моделями **фантастической сложности**.
- МСМС работает **медленно**.
- Генерируемая выборка только **в пределе** похожа на выборку из апостериорного закона.

Конструктор структурных моделей

Компоненты: тренд, сезонность, предикторы, ошибка:

$$y_t = \mu_t + s_t + \beta_t x_t + u_t, \quad u_t \sim \mathcal{N}(0; \sigma_{obs}^2).$$

Для каждой компоненты есть куча вариантов.

Тренд

$$y_t = \mu_t + s_t + \beta_t x_t + u_t, \quad u_t \sim \mathcal{N}(0; \sigma_{obs}^2).$$

Локальный линейный тренд:

$$\mu_t = \mu_{t-1} + \delta_{t-1} + w_{1t}, \quad w_{1t} \sim \mathcal{N}(0; \sigma_{level}^2).$$

Уравнение для наклона δ_t :

$$\delta_t = \delta_{t-1} + w_{2t}, \quad w_{2t} \sim \mathcal{N}(0; \sigma_{slope}^2).$$

Сезонность с помощью дамми

$$y_t = \mu_t + s_t + \beta_t x_t + u_t, \quad u_t \sim \mathcal{N}(0; \sigma_{obs}^2).$$

γ_{it} — оценка сезонного эффекта для наблюдения $t - i$ в момент t .

$$s_t = \gamma_{0t},$$

$$\gamma_{it} = \gamma^{i-1, t-1}, \quad i \in \{1, \dots, 11\}. \gamma_{0t} + \gamma_{1, t-1} + \gamma_{2, t-1} + \dots$$

Сезонность с помощью Фурье

$$y_t = \mu_t + s_t + \beta_t x_t + u_t, \quad u_t \sim \mathcal{N}(0; \sigma_{obs}^2).$$

$$s_t = a_{1t} \cos\left(\frac{2\pi}{365}t\right) + b_{1t} \sin\left(\frac{2\pi}{365}t\right) + a_{2t} \cos\left(2 \cdot \frac{2\pi}{365}t\right) + b_{2t} \sin\left(2 \cdot \frac{2\pi}{365}t\right)$$

$$a_{it} = a_{i,t-1} + w_{4it}, \quad w_{4it} \sim \mathcal{N}(0; \sigma_{ai}^2) \quad b_{it} = b_{i,t-1} + w_{5it}, \quad w_{5it} \sim \mathcal{N}(0; \sigma_{bi}^2)$$

Эволюция зависимости от предиктора

$$y_t = \mu_t + s_t + \beta_t x_t + u_t, \quad u_t \sim \mathcal{N}(0; \sigma_{obs}^2).$$

$$\beta_t = \beta_{t-1} + w_{6t}, \quad w_{6t} \sim \mathcal{N}(0; \sigma_{reg}^2)$$

Байесовские структурные модели: итоги

- Трактуем параметры как **случайные величины**.
- Конструируем из кубиков **настоящего монстра** с предикторами и **кучей** параметров.
- MCMC работает **медленно**.
- С помощью MCMC можно сгенерировать большую выборку из **апостериорного** распределения параметров.

Как оценить эффект воздействия?

Оценивание эффекта: план

- Основная идея оценивания.
- Чем хорош байесовский подход?

Оценивание эффекта: идея

- Возьмём **любую** модель.
- Разделим выборку на обучающую и тестовую **по точке воздействия**.
- Оценим модель по обучающей выборке и **получим прогноз**.
- **Разница** прогноза и фактических значений на тестовой выборке оценивает эффект воздействия.

Нюансы идеи

- Модель должна **хорошо** прогнозировать.
Если сезонность сильна, то даже **наивная сезонная** модель подойдёт.
- Разумно **протестировать идею** в точке, где воздействия ещё нет.
- Можно оценивать эффект воздействия на конкретное значение y_{T+h} , а можно оценивать **кумулятивный эффект** на $y_{T+1} + y_{T+2} + \dots + y_{T+h}$.
- Помните о **доверительных** интервалах.

А зачем тут байесовский подход?

- Позволяет оценивать **более сложные** модели с предикторами.

Возможность более точно оценить эффект воздействия.

- Позволяет построить **доверительный интервал** для кумулятивного эффекта.

Проблема: прогнозы $\hat{y}_{T+1}, \dots, \hat{y}_{T+h}$ нетривиально коррелированы между собой. В частотном подходе сложно получить **явную формулу** для доверительного интервала.

Решение: в байесовском подходе апостериорная выборка параметров модели позволяет сгенерировать множество **гипотетических будущих** траекторий без воздействия.

- Графики **будущих** траекторий!

Оценивание эффекта воздействия: итоги

- Делим выборку на обучающую и тестовую **в точке воздействия**.
- Смотрим на **разницу** прогноза модели и фактических значений.
- Байесовский подход дарит **доверительные интервалы** для кумулятивного эффекта.

Неужели на том и конец?

Спасибо!