3 Анализ временных рядов

3.1 Введение

Многие задачи науки и техники связаны с процессами $X(t)$, которые можно представить в виде совокупности измерений x_t на некотором интервале времени. Значения процесса x_t в каждый момент времени t является случайной величиной. Такие процессы называют временными рядами.

Временным рядом назовем (динамическим рядом) назовем последовательность наблюдений некоторого признака (случайной величины) X в последовательные (как правило, равноотстоящие) моменты времени t. Отдельные наблюдения называются уровнями ряда, обозначим их x_t $(t = 1, 2, ..., n)$, где n – число уровней. Раньше вариационный ряд x_1, x_2, ..., x_n рассматривался как одна из реализаций случайной величины X, временной ряд x_1, x_2, ..., x_n можно рассматривать как одну из реализаций случайного процесса $X(t)$. Однако, имеется принципиальное различие между временным рядом x_t $(t = 1, 2, ..., n)$ и последовательностью наблюдений x_1, x_2, ..., x_n, образующих случайную выборку. В отличие от элементов выборки члены временного ряда, как правило, не являются статистически независимыми и одинаково распределенными.

Примером временного ряда может служить переменная x_t, полученная при наложении случайных флуктуаций на детерминированную (неслучайную) составляющую, которую называют *трендом* временного ряда.

В тех случаях, когда измерения могут регистрироваться (по крайней мере, теоретически) непрерывно, соответствующие временные ряды называют непрерывными. Если временной параметр t является дискретным, мы имеем дискретные временные ряды.

Для временных рядов основной интерес представляет моделирование их структуры с дальнейшим применением модели для экстраполяции или прогнозирования. При исследовании временных рядов необходим статистический подход из-за ошибок измерений и случайных флуктуаций, свойственных практически любой наблюдаемой системе, относится ли она к медицине, биологии, окружающей среде или технике. Важной составляю- щей статистических методов анализа временных рядов является оценка тренда, который, при наличии информации об его виде, можно моделировать при помощи компонент, являющихся детерминированными функциями времени. Отметим, что эксперименты, в ко- торых осуществляются наблюдения, как правило, не являются независимыми, и последовательные ошибки модели должны, вообще говоря, рассматриваться как статистически связанные. Обычной практикой является предположение, что временной ряд наблюдается в равноотстоящие друг от друга моменты времени t_1, t_2, ..., t_n и что последовательные ошибки модели образуют стационарный временной ряд, вероятностные свойства которого не изменяются во времени.

Временной ряд x_t $(t = 1, 2, ..., n)$ называется строго стационарным, если совместное распределение вероятностей n наблюдений x_1, x_2, ..., x_n такое же, как и n наблюдений $x_{1+\tau}, x_{2+\tau}, ..., x_{n+\tau}$ при любых n, t и τ. То есть, свойства строго стационарных временных рядов (закон распределения и числовые характеристики) не зависят от t. Следовательно, математическое ожидание и дисперсия могут быть оценены по наблюдениям x_t $(t = 1, 2, ..., n)$ как мы это делали ранее через $\bar{X}$ и $\bar{S}^2$.

Моделью некоторых нестационарных временных рядов служат процессы вида $y_t = \mu(t) + x_t$ где детерминированная функция $\mu(t)$ зависит лишь от t, а x_t – стационарный временной ряд с нулевым средним. Детерминированная компонента $\mu(t)$ характеризует тенденцию изменения временного ряда y_t в среднем со временем, т.е. его тренд, а слагаемое x_t определяет случайные не зависящие от t ошибки модели и структуру их зависимости.

3.2 Сглаживание временного ряда

Как уже говорилось выше, одной из задач теории временных рядов является оценивание тренда временного ряда x_t – неслучайной составляющей $\mu(t)$ по результатам наблюдений x_1, x_2, ..., x_n его отрезка длины n. Для решения этой задачи необходимо выбрать вид функции $\mu(t)$. Наиболее часто используются следующие функции:

1. Линейная: $\mu(t) = a_0 + a_1 t$.
2. Полиномиальная: $\mu(t) = a_0 + a_1 t + a_2 t^2 + ... + a_r t^r$.
3. Экспоненциальная: $\mu(t) = e^{a_0 + a_1 t}$.
4. Логистическая: $\mu(t) = \frac{a_0}{1 + a_1 e^{-a_2 t}}$.

Оценивание тренда временного ряда можно осуществить при помощи операции сглаживания, целью которой является уменьшение дисперсии временного ряда x_t, а, по существу, амплитуды случайных флуктуаций вокруг его детерминированной составляющей.

Например, для сглаживания можно выбрать линейную функцию $\mu(t) = \hat{\beta}_0 + \hat{\beta}_1 t$, где $\hat{\beta}_0$ и $\hat{\beta}_1$ вычисляется методом наименьших квадратов.

$$\begin{cases} \hat{\beta}_1 = \dfrac{\sum\limits_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sum\limits_{i=1}^{n}(x_i - \bar{X})^2}, \\ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}. \end{cases} \tag{53}$$

Либо, например, можно выбрать квадратичную функцию $\mu(t) = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2$ для которой МНК даёт

$$\begin{cases} \beta_2 = \dfrac{S_{xx}S_{xxy} - S_{xxx}S_{xy}}{S_{xx}S_{xxxx} - S_{xxx}^2 - \frac{S_{xx}^3}{n}}, \\ \beta_1 = \dfrac{S_{xy} - \beta_2 S_{xxx}}{S_{xx}} - 2\beta_2 \bar{X}, \\ \beta_0 = \bar{Y} - \beta_1 \bar{X} - \beta_2 \bar{X^2}, \end{cases} \tag{54}$$

где

$$\begin{aligned} &\bar{X^2} = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} x_i^2, S_{xx} = \sum\limits_{i=1}^{n}(x_i - \bar{X})^2, S_{xxx} = \sum\limits_{i=1}^{n}(x_i - \bar{X})^3, \\ &S_{xxxx} = \sum\limits_{i=1}^{n}(x_i - \bar{X})^4, S_{xy} = \sum\limits_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y}), \\ &S_{xxy} = \sum\limits_{i=1}^{n}(x_i - \bar{X})^2(y_i - \bar{Y}). \end{aligned} \tag{55}$$

Другим методом выравнивания временного ряда является метод экспоненциального сглаживания или его частный случаем – *метод скользящего среднего*. При любом использованном методе сглаживания исследователь рассчитывает, что полученный в результате сглаживания новый временной ряд y_t будет иметь более четко выраженный тренд, мало отличающийся от тренда первоначального ряда x_t и, следовательно, в первом приближении могущий его заменить. Приведем формулу для метода скользящего среднего по трем точкам:

$$y_t = \frac{1}{3}(x_{t-1} + x_t + x_{t+1}), \tag{56}$$

где $2 \le t \le n - 1$. Этот метод сглаживания оставляет линейный тренд временного ряда x_t без изменения и, вообще говоря, уменьшают его дисперсию.

Отметим также, что при сглаживании, как правило, происходит некоторая потеря информации. Так, при использовании метода скользящего среднего по трём точкам отрезок сглаженного ряда будет содержать $n-2$ элемента вместо n.

Знание тренда $\mu(t)$ временного ряда позволяет с известной степенью надежности и точности прогнозировать его значения. Предположительные значения ряда для моментов времени τ, выходящих за границу проведенных наблюдений, считаются равными $\mu(t)$ При этом важно понимать, что оценки подобного сорта предполагают то, что основная тенденция изменения временного ряда в течение интервала времени между моментом наблюдения и моментом времени, для которого оценивается значение временного ряда, сохраняется. Кроме того, надо иметь в виду, что точность прогноза, как правило, снижается с ростом этого интервала.

Пример 3.7. Требуется дать прогноз объема производства в 2009 г., применяя сглаживание временного ряда с помощью как линейной, так и квадратичной функции. Линейная аппроксимация: $\hat{y}_i(x) = 5.28x_i + 8.84$. Прогноз $\hat{y}_i(2009) = 24.15$. Квадратичная аппроксимация: $\hat{y}_i(x) = 1.73x_i^2 + 5.28x_i + 5.38$. Прогноз $\hat{y}_i(2009) = 35.23$. При этом, из графика на рисунке 7 можно сделать вывод, что прогноз по квадратичной аппроксимации точнее линейного.

Год	1960	1970	1980	1990	2000	2009
Млн. т.	1.6	2.1	5.4	12.1	23.0	?

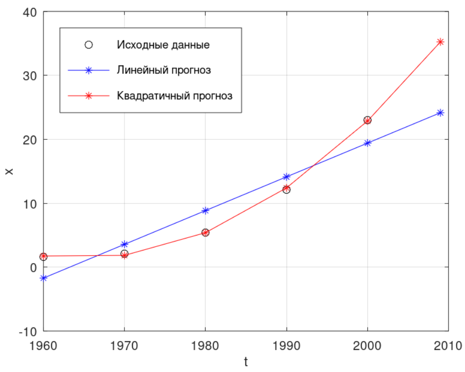

Рис. 7.: Линейный и квадратичный тренды

3.3 Ход выполнения работы

На языке программирования Python для заданных данных:

1. Для полученного временного ряда получить прогноз, применяя выявление тренда временного ряда методом наименьших квадратов с помощью линейной и квадратичной функции.
2. На одном графике построить изображение ряда, а также линейного и квадратичного трендов.
3. Для выполнения работы разрешается использовать базовые вычислительные операции библиотек *numpy* или *math*, такие как сложение, вычитание, деление и т.д.