**Прогнозирование с помощью временных рядов.**

**Экспоненциальное среднее как предиктор.**

Определение: Предиктор - модель, служащая для предсказания.

1. Постановка задачи прогнозирования.

Проблему прогнозирования можно определить как задачу оценивания по данной последовательности чисел, взятых из какого либо временного ряда последующих значений того же ряда.

Пусть в последовательности дискретных наблюдений каждое значение представляет собой сумму :

хt=аt+t (1)

t - момент времени измерения t=1,2,..

аt – уровень ряда (модель ряда),

t -случайная помеха (“белый шум”), имеющая нормальный закон распределения с М(t )=0 и D(t )=.

Будем задавать аt в виде полинома n-го порядка.

 

Задача предсказания значения, отстоящего на m шагов от последнего наблюдавшегося значения xt  включает следующие этапы:

1) Выбор дискретности наблюдения Δt и интервала наблюдения Т.

2) Выбор модели процесса аt, т.е. определение порядка полинома в формуле (2).

3) Вычисление оценок коэффициентов модели по заданным значениям ряда, наблюдаемым на интервале Т.

4) Использование полученной модели для предсказания значений 

5) Оценивание точности предсказания.

Основной этап задачи прогнозирования - этап сглаживания.

В случае, когда веса наблюдений убывают по экспоненте, сглаживание называется экспоненциальным.

Если n=0, то a=const.

**Многократное экспоненциальное сглаживание.**

Оператор простого экспоненциального сглаживания – это оператор экспоненциального сглаживания первого порядка. Если к результату простого экспоненциального сглаживания вновь применить, ту же процедуру, то получим оператор сглаживания второго порядка:





- оператор экспоненциального сглаживания третьего порядка.

...

- оператор экспоненциального сглаживания n-го порядка.

Основная теорема экспоненциального сглаживания.

Обозначим прогнозируемое значение через . Его можно выразить с помощью ряда Тейлора для наблюдавшегося значения в момент времени t.

, где (3)



m - количество интервалов от последнего наблюдавшегося значения хt до предсказываемого значения .

Из сравнения формул (2)и (3) можно сделать вывод, что неизвестный коэффициент i представляет собой соответствующие производные переменной х в момент времени t. Значение производных можно определить, пользуясь сглаженными величинами.

Основная теорема экспоненциального сглаживания позволяет получить оценки n+1 коэффициента (производной) полиномиальной модели n-го порядка как линейную комбинацию результата первых n+1 порядков сглаживания.

**Теорема:** Если ряд наблюдений xt представлен моделью , то существует система n+1 уравнений, связывающих сглаженные величины  с производными , т.е.



Решение этой системы дает выражения для производных в виде линейных комбинаций сглаженных данных.

Применяя основную теорему для случая, когда моделью является постоянная величина ξ = а, т. е. n = 0, получаем всего одно уравнение:

. (4)

Основная теорема позволяет получить искомые оценки коэффициентов А*i*– по следующей схеме: наблюдавшиеся данные — сглаженные величины — оценки коэффициентов.

Вывод расчетных формул для коэффициентов линейной модели .

аt=А0(t)+А1(t) \*t, (5)

t - номер наблюдения

шаг 1. Пользуясь определением многократного сглаживания, выразим текущие сглаженные величины через текущие наблюдавшиеся значения и предыдущие сглаженные величины:



шаг 2. Выразим коэффициенты модели через сглаженные величины. Для этого необходимо на основании теоремы экспоненциального сглаживания записать выражение для сглаженных величин через оценки коэффициентов А0 и А1 .

(\*\*)

Решив эту систему относительно А0(t) и А1(t), получим:





(\*\*\*)

шаг 3. Подставив выражение для  и  из выражений (\*) и (\*\*\*) получим:



шаг 4. По аналогии с уравнениями (\*\*) запишем выражения для предыдущих сглаженных величин через оценки коэффициентов моделей:



шаг 5. Подставим выражение для  и  из формул пункта 4 в формулы пункта 3.

**Окончательное выражение для оценок коэффициентов линейной модели:**

, где (6)

**Выбор дискретности отсчетов**.

Интервал между замерами рекомендуется выбирать в зависимости от требуемого минимального времени упреждения *τ = m∆t.* Обычно интервал между замерами *∆t* берут равным 0,01—0,25 минимального времени упреждения. Указанная рекомендация не является жесткой, однако следует помнить, что при слишком малом интервале между замерами характер процесса затемняется шумом. С другой стороны, при увеличении *∆t* растет величина ошибки предсказанного значения.

**Выбор модели**.

Для определения порядка полинома на основе выборки вычисляют разности первого порядка *∆x*(t) *= x*t*+*1 *—* xt*;* если они колеблются около нуля, то данные можно аппроксимировать постоянной величиной. Если среднее разностей первого порядка отлично от нуля, а для разностей второго порядка *∆*(2)*x* = *∆x*(t+1)- *∆x*(t)оно равно нулю, то данные представляют с помощью линейного закона. Вообще, если среднее разностей *(п-*1)-го порядка отлично от нуля, а среднее разностей n-го порядка равно нулю, то моделью служит полином (n—1)-й степени, причем среднее значение разности (n —1)-го порядка может служить начальной оценкой коэффициента при старшем члене полинома. Если наблюдается систематический рост разностей, то возможно, что данные описываются экспонентой. Для проверки необходимо оценить отношение двух соседних наблюдений. При экспоненциальном законе процентный рост данных должен быть постоянным.

При подборе модели не следует стремиться описывать весь ряд целиком: модель должна хорошо представлять лишь некоторый отрезок, который простирается достаточно далеко в прошлое (чтобы иметь достаточное количество данных для оценивания коэффициентов модели) и в будущее (чтобы охватить интервал упреждения). Этот отрезок перемещается вдоль временного ряда по мере того, как поступают новые данные.

Выбор начальных условий.

Т.к. оператор сглаживания требует рекуррентных вычислений, то для оценивания коэффициентов модели необходимо знать начальные величины.



Обычно удобнее оценивать не начальное значение сглаживаемых величин, а начальные значения коэффициентов ряда Тейлора, т.е. 

Эти начальные оценки можно получить на основе прошлых данных, например для постоянной модели —путем усреднения, для линейной модели — методом наименьших квадратов. Также для определения начальных значений коэффициентов линейной модели можно построить график выборки. Зная оценки  ии пользуясь уравнениями , (7)

можно вычислить начальные оценки для сглаженных величин:

, (8)

**Прогнозирование**. После того как получены оценки коэффициентов модели, может быть вычислена оценка будущего наблюдения:

, (9)

где *п —* порядок полинома;  - оценка коэффициента Аi для наблюдения с номером t; m - количество интервалов от последнего наблюдавшегося значения хt до предсказанного .

**Оценка точности предсказания.**

Для любой последовательности данных полином n-ой степени полученный путем многократного экспоненциального сглаживания является решением минимизирующим взвешенную сумму квадратов разностей:



При условии отсутствия шума многократное экспоненциальное сглаживание обеспечивает получение точных коэффициентов моделей. Обычно данные наблюдения включают в себя шум. Благодаря линейности оператора сглаживания можно записать:

St(x) = St(a) + St()

Если М(t)=0, то М(St(x))= М(St(а)).

Оценки коэффициентов моделей, полученные в результате многократного сглаживания, имеют нормальное распределение.

Будем вычислять оценку дисперсии  по экспериментальным данным на основании сравнения предсказанных величин и соответствующих им действительно наблюдавшихся значений. (n - порядок полинома)

 (10)

**Задание и порядок выполнения работы.**

1) Загрузить *N* значений временного ряда, соответственно варианту (Приложение 2). Для данной выборки найти оценки математического ожидания и дисперсии. Построить график , *t=1,2,…,N* полученных значений временного ряда.

2)Использовать постоянную модель (n=0) для прогнозирования значений временного ряда при  и .

 считать прогнозом на момент времени t+1. Начальное значение  найти как среднее арифметическое пяти первых значений временного ряда. На каждом шаге сравнить предсказанное и действительное значения, выписывая ошибку предсказания -, *t=1,2,…,N.* Вычислить выборочную дисперсию  (10) ошибки предсказания по данным эксперимента.

3) Оценить начальные значения коэффициентов линейной модели и, воспользовавшись методом наименьших квадратов (см. Приложение 1), то есть по значениям  нужно построить аппроксимирующую прямую. Точка пересечения этой линии с осью ординат определяет величину , а тангенс угла наклона к оси абсцисс – величину .

4) Используя операторы экспоненциального сглаживания первого и второго порядков (пункт «Многократное экспоненциальное сглаживание»), получить сглаженные значения  и , *t=1,2,…,N* при значениях постоянной сглаживания  и . Построить графики (всего четыре графика). (Начальные значения найти по формуле (7) или (8)).

5) Использовать линейную модель (n=1)для прогнозирования значений временного ряда *х* с интервалом упреждения

*τ =∆t*, что равносильно m=1 при  и . (Формулы (6),(9)).

На каждом шаге сравнить предсказанное и действительное значения, выписывая ошибку предсказания -, *t=1,2,…,N.* Вычислить выборочную дисперсию ошибки предсказания по данным эксперимента 

6) Выполнить пункт 5) для m=5. Сравнить полученные результаты.

7) Представить рекомендацию по выбору модели (порядка полинома) для заданного временного ряда по приведенному в пункте «Выбор модели» алгоритму.

Использованная литература

1. Статистические методы в инженерных исследованиях под ред. Г.К.Круга (Лабораторный практикум).

Приложение 1.

## Простая линейная регрессия

Имеются экспериментальные значения случайных величин x и y: (,(,…,(.

Будем искать зависимость вида: y = a0 + a1 x







Приложение 2.

Приведены данные продаж на конец недели. Данные записывались в течение 30 недель.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| вариант | 1, 20 | 2, 21 | 3, 22 | 4, 23 | 5, 24 | 6, 25 | 7, 26 | 8, 27 | 9, 28 | 10, 29 | 11, 30 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| неделя |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 | 33 | 43 | 240 | 272 | 1795 | 1523 | 1630 | 622 | 666 | 53 | 72 | 517 | 609 | 1204 | 1494 | 6827 | 6848 | 737 | 2428 |
| 2 | 35 | 44 | 260 | 281 | 1738 | 1549 | 1659 | 620 | 670 | 33 | 73 | 538 | 625 | 1328 | 1525 | 6178 | 7027 | 775 | 2010 |
| 3 | 37 | 45 | 265 | 289 | 1934 | 1576 | 1689 | 621 | 676 | 30 | 77 | 554 | 644 | 1328 | 1551 | 7084 | 7685 | 792 | 2981 |
| 4 | 40 | 42 | 267 | 291 | 1835 | 1602 | 1720 | 630 | 684 | 29 | 81 | 575 | 665 | 1435 | 1539 | 8162 | 7602 | 787 | 3074 |
| 5 | 38 | 47 | 269 | 296 | 2024 | 1630 | 1749 | 636 | 696 | 55 | 78 | 584 | 676 | 1416 | 1629 | 8462 | 7775 | 835 | 2893 |
| 6 | 43 | 48 | 268 | 299 | 2083 | 1659 | 1778 | 650 | 705 | 44 | 79 | 601 | 700 | 1494 | 1665 | 9644 | 7933 | 887 | 3198 |
| 7 | 44 | 49 | 272 | 302 | 1341 | 1689 | 1807 | 666 | 707 | 41 | 87 | 609 | 725 | 1525 | 1708 | 8350 | 8094 | 810 | 3250 |
| 8 | 45 | 53 | 281 | 306 | 987 | 1720 | 1837 | 670 | 718 | 43 | 94 | 625 | 745 | 1551 | 1799 | 7829 | 9280 | 832 | 3495 |
| 9 | 42 | 56 | 289 | 310 | 1650 | 1749 | 1865 | 676 | 731 | 68 | 93 | 644 | 787 | 1539 | 1873 | 8829 | 8730 | 855 | 3528 |
| 10 | 47 | 60 | 291 | 315 | 2074 | 1778 | 1892 | 684 | 745 | 55 | 84 | 665 | 810 | 1629 | 1973 | 9948 | 9614 | 878 | 3838 |
| 11 | 48 | 66 | 296 | 324 | 2122 | 1807 | 1919 | 696 | 758 | 55 | 92 | 676 | 832 | 1665 | 2087 | 10638 | 9290 | 884 | 3916 |
| 12 | 49 | 72 | 299 | 334 | 1920 | 1837 | 1943 | 705 | 773 | 67 | 100 | 700 | 855 | 1708 | 2208 | 11253 | 10925 | 913 | 4142 |
| 13 | 53 | 73 | 302 | 348 | 1877 | 1865 | 1966 | 707 | 787 | 55 | 106 | 725 | 878 | 1799 | 2271 | 11179 | 10645 | 941 | 4441 |
| 14 | 56 | 77 | 306 | 367 | 1815 | 1892 | 1987 | 718 | 807 | 57 | 110 | 745 | 884 | 1873 | 2365 | 12820 | 12161 | 959 | 5583 |
| 15 | 60 | 81 | 310 | 388 | 1848 | 1919 | 2007 | 731 | 828 | 52 | 108 | 787 | 913 | 1973 | 2423 | 12950 | 10466 | 939 | 6230 |
| 16 | 66 | 78 | 315 | 405 | 1646 | 1943 | 2027 | 745 | 844 | 34 | 111 | 810 | 941 | 2087 | 2416 | 10894 | 11030 | 957 | 6497 |
| 17 | 72 | 79 | 324 | 418 | 1653 | 1966 | 2051 | 758 | 870 | 29 | 103 | 832 | 959 | 2208 | 2484 | 10455 | 11424 | 983 | 5480 |
| 18 | 73 | 87 | 334 | 444 | 1810 | 1987 | 2077 | 773 | 894 | 30 | 109 | 855 | 939 | 2271 | 2605 | 11179 | 10748 | 1000 | 5870 |
| 19 | 77 | 94 | 348 | 493 | 1462 | 2007 | 2099 | 787 | 920 | 28 | 121 | 878 | 957 | 2365 | 2744 | 10590 | 11390 | 1002 | 6354 |
| 20 | 81 | 93 | 367 | 538 | 1404 | 2027 | 2110 | 807 | 938 | 28 | 110 | 884 | 983 | 2423 | 2729 | 8919 | 11637 | 996 | 6610 |
| 21 | 78 | 84 | 388 | 569 | 1522 | 2051 | 2138 | 828 | 962 | 41 | 115 | 913 | 1000 | 2416 | 2695 | 11607 | 12200 | 993 | 6290 |
| 22 | 79 | 92 | 405 | 606 | 1624 | 2077 | 2160 | 844 | 990 | 50 | 125 | 941 | 1002 | 2484 | 2826 | 12537 | 11577 | 1007 | 6725 |
| 23 | 87 | 100 | 418 | 652 | 1732 | 2099 | 2180 | 870 | 1020 | 49 | 145 | 959 | 996 | 2605 | 2858 | 14759 | 12246 | 1003 | 6435 |
| 24 | 94 | 106 | 444 | 726 | 1850 | 2110 | 2202 | 894 | 1050 | 44 | 132 | 939 | 993 | 2744 | 3115 | 10437 | 13281 | 1030 | 6687 |
| 25 | 93 | 110 | 493 | 824 | 1920 | 2138 | 2226 | 920 | 1070 | 52 | 136 | 957 | 1007 | 2729 | 3190 | 13589 | 10360 | 1055 | 6885 |
| 26 | 84 | 108 | 538 | 909 | 2074 | 2160 | 2251 | 938 | 1090 | 79 | 158 | 983 | 1003 | 2695 | 3248 | 13402 | 13812 | 1077 | 6540 |
| 27 | 92 | 111 | 569 | 965 | 2122 | 2180 | 2277 | 962 | 1150 | 68 | 146 | 1000 | 1030 | 2826 | 3166 | 13103 | 12185 | 1040 | 6480 |
| 28 | 100 | 103 | 606 | 996 | 2305 | 2202 | 2300 | 990 | 1120 | 83 | 148 | 1002 | 1055 | 2858 | 3279 | 14190 | 14057 | 1280 | 7000 |
| 29 | 106 | 109 | 652 | 103 | 2280 | 2226 | 2323 | 1020 | 1160 | 107 | 160 | 996 | 1077 | 3115 | 3501 | 13560 | 16243 | 1090 | 6580 |
| 30 | 110 | 121 | 726 | 107 | 2295 | 2251 | 2340 | 1050 | 1090 | 105 | 155 | 993 | 1040 | 3190 | 3618 | 10820 | 12400 | 1150 | 6985 |