1 Методология

В данной работе сравниваются три метода на примере задачи экстраполяции таблично заданной функции “недалеко” от границы определения. Под “недалеко” имеется в виду в пределах 10% от диапазона узлов за границей области определения. В качестве функции взята численность населения США в 1900-2020 гг. Экстраполяция проводится по всем годам, кроме последнего, а затем сравнивается вычисленное значение в 2020г с настоящим.

Из трёх используемых методов два являются интерполяционными методами. Несмотря на это, они всё равно могут быть использованы для экстраполяции при выполнении определённых условий (см. “Описание методов”).

Примечание: В методе наименьших квадратов используется полином максимальной возможной степени для данного количества узлов: $n = 11$.

2 Описание методов

В данном разделе описаны три метода приближения значений функций: классическая полиномиальная интерполяция по Ньютону, сплайн-интерполяция и метод наименьших квадратов. Каждый метод представлен с указанием типа, условий применимости, алгоритма и особенностей.

2.1 Классическая полиномиальная интерполяция по Ньютону

- **Тип метода:** Интерполяция (точное прохождение через заданные узлы) с поддержкой экстраполяции.
- **Характеристики:**
  - Полиномиальная (использует алгебраический полином).
  - Точечная (использует таблицу значений (x_i, f_i)).
  - Универсальная (узлы не обязательно равноотстоящие).
  - Форма Ньютона удобна для добавления новых узлов без пересчёта всех коэффициентов.
  - Поддерживает экстраполяцию: позволяет вычислять значения за пределами интервала узлов.

Условия применимости и ограничения

1. Задана таблица из $(n + 1)$ точки (x_i, f_i), где $i = 0, 1, \ldots, n$.
2. Все узлы x_i различны (условие единственности интерполяционного полинома).
3. Функция $f(x)$ предполагается непрерывной и достаточное число раз дифференцируемой на интервале, содержащем узлы (для оценки погрешности).
4. **Экстраполяция:** Метод позволяет вычислять значения за пределами интервала узлов, однако при этом:
   - Погрешность экстраполяции обычно значительно превышает погрешность интерполяции.
   - При удалении от области узлов погрешность растёт быстро (иногда экспоненциально).

- Особенно опасна экстраполяция полиномами высокой степени (более 5-6).
- Рекомендуется экстраполяция “недалеко” за границы области узлов (в пределах 10-20% от диапазона узлов).

Алгоритм метода

Пусть заданы узлы $x_0, x_1, \ldots, x_n$ и значения $f_i = f(x_i)$.

1. **Построение разделённых разностей:**

   - Разделённые разности нулевого порядка: $f[x_i] = f_i$.
   - Разделённые разности первого порядка:

   $$f[x_i, x_{i+1}] = \frac{f[x_{i+1}] - f[x_i]}{x_{i+1} - x_i}.$$

   - Разделённые разности k-го порядка (рекуррентно):

   $$f[x_i, x_{i+1}, \ldots, x_{i+k}] = \frac{f[x_{i+1}, \ldots, x_{i+k}] - f[x_i, \ldots, x_{i+k-1}]}{x_{i+k} - x_i}.$$

2. **Построение интерполяционного полинома Ньютона:**

   $$P_n(x) = f[x_0] + f[x_0, x_1](x-x_0) + f[x_0, x_1, x_2](x-x_0)(x-x_1) + \cdots + f[x_0, x_1, \ldots, x_n](x-x_0)(x-x_1)\ldots(x-x$$

   Или в компактной форме:

   $$P_n(x) = \sum_{k=0}^{n} \left(f[x_0, \ldots, x_k] \cdot \prod_{j=0}^{k-1} (x - x_j) \right).$$

3. **Вычисление значения в произвольной точке x:** Подставить x в построенный полином $P_n(x)$. При этом:

   - Если $x \in [\min(x_i), \max(x_i)]$ — это интерполяция.
   - Если $x < \min(x_i)$ или $x > \max(x_i)$ — это экстраполяция.

Погрешность метода

Для любой точки x (включая область экстраполяции) справедливо:

$$f(x) - P_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \cdot \omega_n(x), \quad \omega_n(x) = \prod_{i=0}^{n} (x - x_i),$$

где ξ — некоторая точка из наименьшего интервала, содержащего все узлы x_i и точку x.

Особенности экстраполяции

- При экстраполяции множитель $\omega_n(x)$ растёт быстрее, чем при интерполяции.
- Для экстраполяции “вперёд” (за x_n) рекомендуется упорядочить узлы по возрастанию и использовать полином в форме Ньютона для интерполяции вперёд.
- Для экстраполяции “назад” (до x_0) можно использовать ту же форму, но более устойчивые результаты даёт переупорядочивание узлов в обратном порядке.

2.2 Сплайн-интерполяция с экстраполяцией

- **Тип метода:** Интерполяционный (значения сплайна совпадают с заданными точками) с поддержкой экстраполяции.
- **Характеристики:**
  - Гладкая аппроксимация (непрерывны функция и её первые две производные).
  - Кубические сплайны (на каждом отрезке используется полином 3-й степени).

Условия применимости

1. Узлы x_i должны быть **строго возрастающими**: $x_0 < x_1 < \cdots < x_n$.
2. Количество точек $n + 1 \geq 2$.
3. Тип граничных условий должен быть определён. В реализации используется **естественный сплайн** ($S''(x_0) = S''(x_n) = 0$).
4. Экстраполяция рекомендуется только вблизи границ интервала ($x \in [x_0 - \delta, x_n + \delta]$ с малым δ).

Алгоритм метода

Пусть заданы узлы $x_0, x_1, \ldots, x_n$ и значения $f_i = f(x_i)$. На каждом отрезке $[x_i, x_{i+1}]$ сплайн имеет вид:

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3.$$

Шаг 1: Вычисление разностей

$$h_i = x_{i+1} - x_i, \quad \Delta_i = \frac{f_{i+1} - f_i}{h_i}, \quad i = 0, \ldots, n - 1.$$

Шаг 2: Решение системы для моментов $M_i = S''(x_i)$
Система для естественного сплайна ($M_0 = M_n = 0$):

$$\begin{cases} \mu_i M_{i-1} + 2M_i + \lambda_i M_{i+1} = d_i, \quad i = 1, \ldots, n - 1, \\ M_0 = M_n = 0, \end{cases}$$

где

$$\mu_i = \frac{h_{i-1}}{h_{i-1} + h_i}, \quad \lambda_i = 1 - \mu_i, \quad d_i = 6\frac{\Delta_i - \Delta_{i-1}}{h_{i-1} + h_i}.$$

Шаг 3: Трёхдиагональная прогонка (метод Томаса)

1. **Прямой ход** (вычисление прогоночных коэффициентов):

$$\alpha_1 = -\frac{\lambda_1}{2}, \quad \beta_1 = \frac{d_1}{2},$$

$$\alpha_i = -\frac{\lambda_i}{2 + \mu_i \alpha_{i-1}}, \quad \beta_i = \frac{d_i - \mu_i \beta_{i-1}}{2 + \mu_i \alpha_{i-1}}, \quad i = 2, \ldots, n - 1.$$

2. **Обратный ход**:

$$M_{n-1} = \beta_{n-1}, \quad M_i = \alpha_i M_{i+1} + \beta_i, \quad i = n - 2, \ldots, 1.$$

Шаг 4: Вычисление коэффициентов сплайна
Для каждого отрезка $i = 0, \dots, n-1$:

$$a_i = f_i, \quad c_i = \frac{M_i}{2},$$

$$d_i = \frac{M_{i+1} - M_i}{6h_i}, \quad b_i = \Delta_i - \frac{h_i(2M_i + M_{i+1})}{6}.$$

Шаг 5: Вычисление значения с экстраполяцией
Для заданного x:

1. Если $x < x_0$, используем первый отрезок: $i = 0$.
2. Если $x > x_n$, используем последний отрезок: $i = n - 1$.
3. Иначе находим индекс i такой, что $x_i \le x \le x_{i+1}$.

Затем вычисляем:

$$S(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3.$$

Особенности экстраполяции

- **Естественный сплайн** имеет нулевую вторую производную на границах, что обеспечивает линейное поведение при экстраполяции на бесконечности.
- На практике кубический сплайн может давать разумные значения при экстраполяции на расстояниях, не превышающих шаг интерполяции.
- Качество экстраполяции ухудшается по мере удаления от границ интервала.
- Рекомендуется использовать экстраполяцию только вблизи границ $[x_0 - h_0, x_n + h_{n-1}]$.

2.3 Метод наименьших квадратов (МНК)

- **Тип метода:** Аппроксимация (сглаживание). МНК не является интерполяцией, так как не требует точного прохождения функции через узловые точки. Он предназначен для построения приближающей функции, *наилучшим образом* описывающей общую тенденцию данных, что особенно полезно при наличии случайных погрешностей в измерениях.
- **Суть метода:** Метод позволяет подобрать функцию заданного вида (например, полином степени m), которая минимизирует сумму квадратов отклонений значений этой функции от заданных значений в узлах.

Условия применимости

- Количество точек n должно быть больше количества искомых параметров $m+1$ (система переопределена).
- Искомая зависимость (вид функции F) должна быть выбрана априори на основе анализа природы данных или задачи (линейная, полиномиальная, экспоненциальная и т.д.).
- Для линейного по параметрам случая ($F(x, \vec{c}) = \sum_{k=0}^{m} c_k \varphi_k(x)$), где $\varphi_k(x)$ — базисные функции (например, x^k), метод сводится к решению *системы нормальных уравнений* или линейной задачи. Для нелинейных моделей требуется применение итерационных методов оптимизации.

Алгоритм (для полиномиальной модели $P_m(x) = c_0 + c_1 x + \cdots + c_m x^m$)

1. **Формулировка задачи.** Даны n пар чисел (x_i, y_i). Требуется найти коэффициенты полинома m-й степени ($m < n-1$), минимизирующие сумму $S = \sum_{i=1}^{n} (y_i - P_m(x_i))^2$.

2. **Составление системы нормальных уравнений.** Необходимое условие минимума S — равенство нулю частных производных $\frac{\partial S}{\partial c_k} = 0$ для $k = 0, \ldots, m$. Это приводит к системе $(m+1)$ линейных уравнений:
$$\sum_{j=0}^{m} c_j \sum_{i=1}^{n} x_i^{j+k} = \sum_{i=1}^{n} y_i x_i^k, \quad k = 0, 1, \ldots, m.$$
В матричной форме: $\mathbf{A}^T \mathbf{A} \vec{c} = \mathbf{A}^T \vec{y}$, где
$$A_{ij} = x_i^{j-1} \quad (i = 1..n, j = 1..m+1), \quad \vec{y} = (y_1, \ldots, y_n)^T.$$

3. **Решение системы.** Решается система относительно вектора коэффициентов $\vec{c} = (c_0, c_1, \ldots, c_m)^T$. Для устойчивости при больших m часто используют не прямое решение нормальных уравнений, а методы на основе QR- или SVD-разложения матрицы $\mathbf{A}$.

4. **Получение аппроксимирующей функции.** Искомая функция — полином $P_m(x)$ с найденными коэффициентами.

* QR-разложение (ортогонально-треугольное) и SVD (сингулярное разложение) — численно устойчивые альтернативы для решения задачи МНК.

3 Исследование

Полученный график представлен на рис. 1. Кривая метода Ньютона ушла в глубоко отрицательные значения и поэтому её значение в узле 2020г. не попало на график.

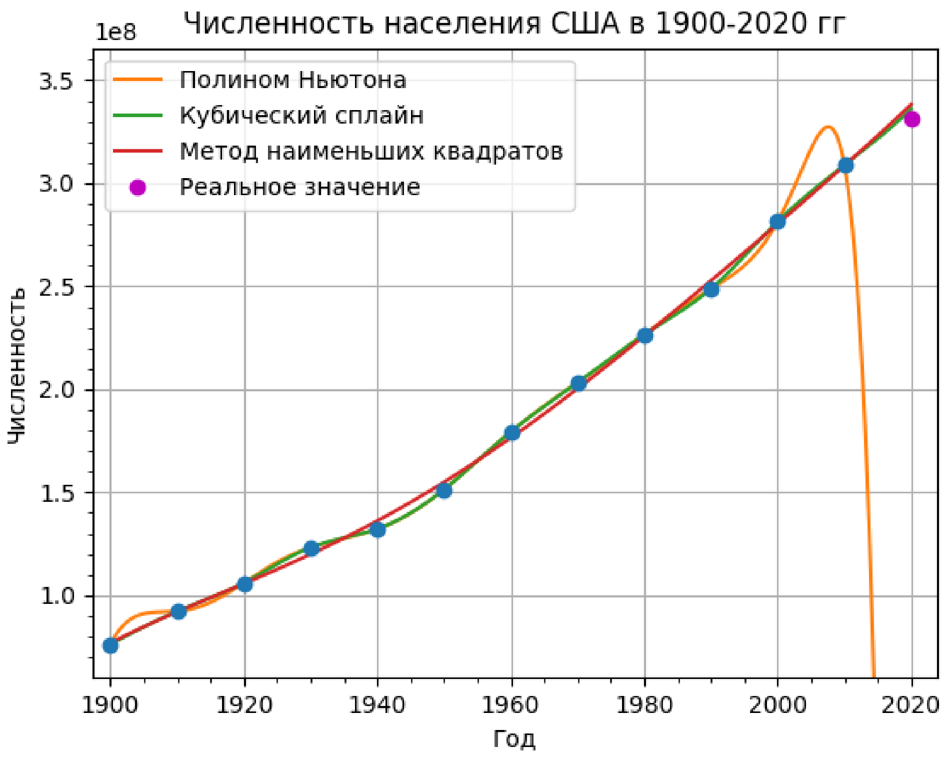

Рис. 1

Полученные относительные погрешности в целевой точке (2020г.):

$$r_{newton} = 545\%$$

$$r_{spline} = 1.4\%$$

$$r_{mnk} = 2.0\%$$

4 Обсуждение результатов

Из полученных результатов очевидно, что метод Ньютона не применим для экстраполяции даже в 10 % от диапазона узлов. Слишком большая степень полинома при 12 узлах привела к выбросу даже между последним и предпоследним узлом.

Метод кубических сплайнов дал наименьшую среди всех методов относительную ошибку экстраполяции несмотря на то, что является интерполяционным методом.