# Глава 7. Восстановление функциональных временных зависимостей

Начнем изложение методов моделирования процессов по временным рядам с наиболее простого подхода – восстановления явной временной зависимости вида $\eta = f(t, \mathbf{c})$, где $f$ – некоторая функция, $\mathbf{c}$ – $P$-мерный вектор параметров модели. Такие задачи рассматриваются в теории аппроксимации функций [22] и математической статистике [155, 171, 1] (см. п. 2.3.1.3).[1] Их можно интерпретировать как проведение кривой через экспериментальные точки на плоскости $(t, \eta)$ или вблизи этих точек (рис.7.1). Умение решать эту задачу в значительной степени определяет успех моделирования и в более сложных ситуациях (см. главы 8-10).

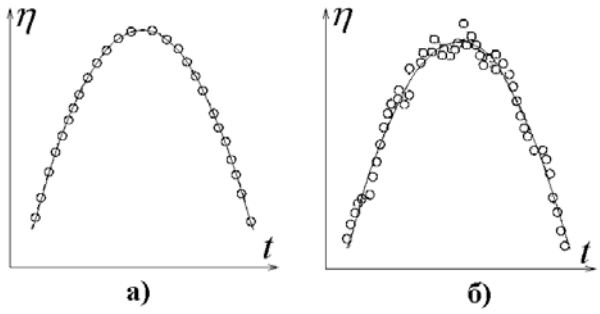



Рис.7.1. Иллюстрация задачи – провести кривую заданного вида через экспериментальные точки на плоскости (кружки). а) Шума нет, вид функции подобрать несложно. б) Точки не ложатся в точности на простую кривую либо из-за случайной помехи, либо из-за сложного характера зависимости. Задача состоит в приближенном описании (аппроксимации) наблюдаемой зависимости простой функцией

Ниже мы рассмотрим эту задачу в двух несколько различных широко распространенных постановках. В первой из них (пп. 7.1, 7.4.2) связь между величинами $t$ и $\eta$ – функция $f(t, \mathbf{c})$ – известна с точностью до значений параметров $\mathbf{c}$. Интерес представляют значения $\mathbf{c}$ потому, например, что они имеют физический смысл, но не могут быть измерены непосредственно. Задача состоит в том, чтобы как можно точнее *оценить параметры* зависимости. Во второй постановке (пп. 7.2, 7.4.1) целью моделирования является возможность предсказать в заданный момент времени $t$ значение $\eta$, т.е. требуется найти функцию $f$, способную обеспечить такой прогноз с минимальной (в определенном смысле) погрешностью. При этом чаще всего вид функции неизвестен – вариант задачи о «черном ящике».

В данной главе вводятся многие важные для дальнейшего изложения идеи и термины. В п. 7.1 рассматриваются и сопоставляются наиболее популярные методы оценки параметров. Затрагиваются вопросы оптимального и устойчивого оценивания. В п. 7.2 вводятся понятия

---

[1] Там обычно говорят о связи величин $X$ и $Y$ произвольной природы. В нашем конкретном случае одной из них является время $t$, другой – наблюдаемая $\eta$.

аппроксимации, регрессии, интерполяции, экстраполяции. Рассматривается выбор класса моделей для аппроксимации и проблема «переобучения» модели. В п. 7.3 обсуждаются вопросы диагностической проверки моделей, в п. 7.4 рассмотрены приложения моделей вида $\eta = f(t, \mathbf{c})$ для прогноза и численного дифференцирования.

## 7.1. Оценка параметров

Рассмотрим задачу о «прозрачном ящике»: структура модели известна, неизвестны конкретные значения параметров. Начнем с детерминированного случая. Исходный процесс задан формулой

$$\eta = f(t, \mathbf{c}_0), \tag{7.1}$$

известен вид функции *f*, неизвестны значения $\mathbf{c}_0$. Задача построения модели вида $\eta = f(t, \mathbf{c})$ сводится к такому подбору значений *P* параметров, чтобы график функции *f* прошел в точности через экспериментальные точки $(t_i, \eta_i), i = 1,..., N$. Решение находится из системы уравнений

$$\eta_i = f(t_i, c_1,..., c_P),\ i = 1,..., n, \tag{7.2}$$

где выбраны *n* моментов наблюдений из общего числа *N*. Если *f* линейна по параметрам, то в типичном случае система (7.2) имеет единственное решение при $n = P$. Оно может быть найдено одним из многих известных методов [153]. Проблема возникает, если матрица системы вырождена или плохо обусловлена (некорректность или плохая обусловленность задачи). Тогда говорят, что параметры модели *неидентифицируемы* по имеющемуся набору данных. Проблема иногда может устраняться при замене одних моментов наблюдений $t_i$ в (7.2) другими. Если число точек *N* во временном ряде больше *P*, то остальные точки могут использоваться для проверки адекватности модели. Если *N* меньше *P*, то задача не имеет единственного решения, т.е. она некорректно поставлена.

В случае нелинейной зависимости *f* от **c** система (7.2) решается приближенными численными методами [73, 85, 153]. Согласно любому из них выбирается некоторая стартовая догадка $\mathbf{c}^{(0)}$ для искомого вектора параметров. Путем решения линеаризованной задачи (метод Ньютона) или иначе получают поправку $\Delta\mathbf{c}^{(0)}$ и новое приближение $\mathbf{c}^{(1)} = \mathbf{c}^{(0)} + \Delta\mathbf{c}^{(0)}$. И так далее, пока итерационный процесс не сойдется с заданной точностью к некоторому решению $\hat{\mathbf{c}}$. При заданной стартовой догадке будет найдено только одно из решений системы, но если она имеет много решений, то нужно найти их все, чтобы потом из них выбрать истинное. Это принципиальная проблема, так как не существует общего метода, который бы гарантировал отыскание всех решений [73], хотя разработано множество более или менее успешных алгоритмов. При небольшом числе

неизвестных задача не очень сложна. Аналогично линейному случаю, достаточно взять $n = P$ (или на 1-2 больше, в зависимости от характера нелинейности).

Рассмотрим подробнее стохастический случай – исходный процесс

$$\eta = f(t, \mathbf{c}_0) + \xi(t), \tag{7.3}$$

где $\xi$ – случайный процесс («шум») с нулевым средним. Здесь речь идет не о точном расчете значений параметров, а о получении их *статистических оценок* $\hat{\mathbf{c}}$, как можно более близких к $\mathbf{c}_0$. Обычно рассматривается ситуация, когда $\xi(t_i)$ – независимые одинаково распределенные случайные величины. Детально закон распределения $\xi$ может быть неизвестен. Тогда либо плотность распределения вероятностей $p(\xi)$ предполагается совпадающей с одним из известных законов (Гаусса, Лапласа, равномерным), либо используются универсальные методы оценки, работоспособные для широкого класса законов распределения.

### 7.1.1. Методы оценивания

Число возможных методов оценивания бесконечно велико – это множество всех функций, которые по входным данным $(t_i, \eta_i), i = 1,..., N$ дают значение $\hat{\mathbf{c}}$ на выходе [81, 145]. Среди этого многообразия есть ряд известных методов, обладающих высокой эффективностью для некоторых классов задач, простотой реализации и т.д. Рассмотрим и сопоставим некоторые из них. В методическом плане полезно различать две ситуации:

1) величина $t$ неслучайна (фиксированные моменты наблюдений),[2] что более типично в задачах анализа временных рядов;
2) величина $t$ случайна (моменты наблюдений выбираются независимо, согласно некоторому закону распределения $p(t)$).

**7.1.1.1. Метод «простого усреднения».**[3] Пусть имеющуюся выборку из $N$ точек можно разделить на $M$ частей по $P$ точек в каждой. Будем действовать так, как если бы шума не было: для каждой $k$-й части решим уравнения (7.2), требующие точного равенства $\eta_i = f(t_i, \mathbf{c})$. Пусть для каждой части они имеют единственное решение $\hat{\mathbf{c}}_k$. Найдем итоговую оценку как эмпирическое среднее $\hat{\mathbf{c}}_k$ (простым усреднением): $\hat{\mathbf{c}} = (1/M)\sum_{k=1}^{M} \hat{\mathbf{c}}_k$ .

[2] Эта ситуация имеет специальное название: схема Гаусса – Маркова [51].
[3] Название не является универсальным, но идея очень проста и часто используется.

**7.1.1.2. Метод статистических моментов.** Метод относится, строго говоря, только к ситуации, когда величина $t$ является случайной, а $f(t,\mathbf{c})$ – алгебраический многочлен. Опишем его на примере исходного процесса

$$\eta = c_1 + c_2 t + c_3 t^2 + \xi(t). \tag{7.4}$$

Найдем математическое ожидание обеих частей равенства (7.4). Используя договоренность $M\xi = 0$, получим [4]

$$M[\eta] = c_1 + c_2 M[t] + c_3 M[t^2]. \tag{7.5}$$

Теперь умножим обе части (7.4) на $\eta$. Опять вычислим среднее и получим

$$M[\eta^2] = c_1 M[\eta] + c_2 M[\eta t] + c_3 M[\eta t^2]. \tag{7.6}$$

Умножим обе части (7.4) на $t$, вычислим среднее, получим

$$M[\eta t] = c_1 M[t] + c_2 M[t^2] + c_3 M[t^3]. \tag{7.7}$$

Если бы были известны значения моментов $M[\eta]$, $M[\eta t]$ и других, входящих в уравнения (7.5) – (7.7), то решая эти линейные уравнения относительно параметров $c_1$, $c_2$, $c_3$, можно было бы точно определить истинные значения последних. Теоретические моменты неизвестны, но заменив их на оценки – эмпирические моменты (пп. 2.2.1.4, 2.2.1.6), т.е. подставив $\langle\eta\rangle = \frac{1}{N}\sum_{i=1}^{N}\eta_i$ вместо $M[\eta]$ и т.д., получим систему уравнений:

$$\begin{aligned} &c_1 + c_2\langle t\rangle + c_3\langle t^2\rangle = \langle\eta\rangle, \\ &c_1\langle\eta\rangle + c_2\langle\eta t\rangle + c_3\langle\eta t^2\rangle = \langle\eta^2\rangle, \\ &c_1\langle t\rangle + c_2\langle t^2\rangle + c_3\langle t^3\rangle = \langle\eta t\rangle, \end{aligned} \tag{7.8}$$

из которой и найдем оценки $\hat{\mathbf{c}}$.

Метод можно обобщить на случай, когда $f$ – не алгебраический многочлен. Значения параметров придется искать, решая нелинейные уравнения, в которые входят величины типа $\langle f(t,\mathbf{c})\rangle$, а не эмпирические моменты. Но практически это очень затруднительно.

Наконец, заметим, что если $t$ – не случайная величина, то интерпретация величин $\langle t\rangle$, $\langle t^2\rangle$ и пр. как оценок статистических моментов

[4] Величина $t$ должна иметь также необходимое число конечных моментов. Так, если $p(t)$ имеет «тяжелые» хвосты (убывает по степенному закону при $t \to \infty$), уже второй эмпирический момент может неограниченно возрастать при увеличении объема выборки.

невозможна. Однако метод применять можно, только следует помнить, что название «метод моментов» не вполне соответствует ситуации.

**7.1.1.3. Метод максимального правдоподобия.** В отличие от двух предыдущих методов, для применения ММП необходимо знать закон распределения[5] $\xi$ с точностью до параметров. ММП был описан в п. 2.2.1.7 в приложении к оценке параметров одномерного распределения. Здесь все обстоит аналогично. Наблюдаемый ряд представляет собой случайный вектор $\{\eta_1,...,\eta_N\}$. При неслучайных моментах $t_i$ и независимых друг от друга $\xi_i$ функция правдоподобия принимает вид

$$L(\mathbf{c}) = p(\eta_1,...,\eta_N|\mathbf{c}) = \prod_{i=1}^{N} p(\eta_i - f(t_i,\mathbf{c})). \tag{7.9}$$

Она зависит от параметров c. ММП реализуется путем максимизации

$$\ln L(\mathbf{c}) = \sum_{i=1}^{N} \ln p(\eta_i - f(t_i,\mathbf{c})) \to \max . \tag{7.10}$$

ММП состоит в том, чтобы выбрать значения параметров **c** так, чтобы плотность вероятности появления наблюдаемого ряда была максимальна.

Если величина $t$ случайна, то практически ничего не меняется: каждый сомножитель (7.9) умножается еще на $p(t_i)$, и, при условии независимости распределения $p(t)$ от **c**, МП-оценки находятся вновь из (7.10).

Конкретный вид функции правдоподобия зависит от закона распределения $p(\xi)$. В общем случае это задача многомерной нелинейной оптимизации, которую решают итерационными методами [73]. Для трех очень распространенных законов распределения $\xi$ ММП приводится к одному из трех других методов: метод наименьших квадратов, метод наименьших модулей, метод минимизации наибольшего уклонения.

**7.1.1.4. Метод наименьших квадратов.** Если $\xi$ распределена по Гауссу $p(\xi) = \left(1\big/\sqrt{2\pi\sigma_\xi^2}\right)\exp\left(-\xi^2\big/2\sigma_\xi^2\right)$ (уместно для описания погрешностей измерений в стабильных условиях [51]), то функция правдоподобия есть

$$\ln L(\mathbf{c}) = -\frac{N\ln(2\pi\sigma_\xi^2)}{2} - \frac{1}{2\sigma_\xi^2}\sum_{i=1}^{N}(\eta_i - f(t_i,\mathbf{c}))^2 . \tag{7.11}$$

Ее максимизация по **c** эквивалентна минимизации величины:

---

[5] Обычно предполагается, что плотность распределения унимодальная.

$$S(\mathbf{c}) = \sum_{i=1}^{N} (\eta_i - f(t_i, \mathbf{c}))^2 \to \min . \qquad (7.12).$$

Это и есть метод наименьших квадратов (МНК). Его геометрический смысл: оценки параметров подбираются так, чтобы минимизировать сумму *квадратов вертикальных расстояний* экспериментальных точек на плоскости $(t, \eta)$ от графика $f(t, \mathbf{c})$, рис.7.2.

**7.1.1.5. Метод наименьших модулей.** Если $\xi$ распределена по Лапласу $p(\xi) = (1/2\Delta)\exp(-|\xi|/\Delta)$ (уместно для описания погрешностей измерений в нестабильных условиях [51]), то функция правдоподобия имеет вид

$$\ln L(\mathbf{c}) = -N \ln(2\Delta) - \frac{1}{\Delta} \sum_{i=1}^{N} |\eta_i - f(t_i, \mathbf{c})| . \quad (7.13)$$

Максимизация по **c** эквивалентна минимизации суммы *модулей вертикальных расстояний*

$$S'(\mathbf{c}) = \sum_{i=1}^{N} |\eta_i - f(t_i, \mathbf{c})| \to \min . \qquad (7.14).$$

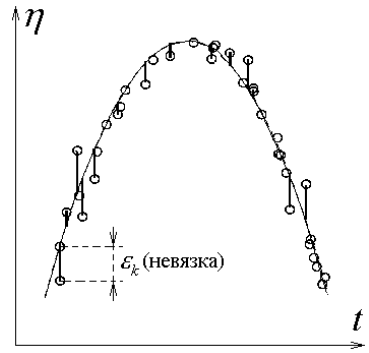



Рис.7.2. Иллюстрация оценки параметров с помощью минимизации вертикальных расстояний (невязок, некоторые показаны черными отрезками) от экспериментальных точек до графика модельной функции. Минимизируется или сумма их квадратов, или сумма модулей, или

**7.1.1.6. Метод минимизации наибольшего уклонения.** Если величина $\xi$ распределена равномерно на отрезке длиной $\Delta$: $p(\xi) = \dfrac{1}{2\Delta}, -\Delta \le \xi \le \Delta$ (уместно для описания погрешностей округления при вычислениях), то функция правдоподобия есть

$$L(\mathbf{c}) = \frac{1}{(2\Delta)^N} \prod_{i=1}^{N} \Theta\left(\Delta - |\eta_i - f(t_i, \mathbf{c})|\right). \quad (7.15)$$

Ее максимизация по **c** (при неизвестном $\Delta$) эквивалентна минимизации наибольшего вертикального расстояния

$$S''(\mathbf{c}) = \max_{1 \le i \le N} |\eta_i - f(t_i, \mathbf{c})| \to \min . \qquad (7.16).$$

Каждый из методов (7.12), (7.14), (7.16) совпадает с ММП только при указанных законах распределения $\xi$. Каждый из них при некоторых условиях превосходит остальные и потому имеет ценность и как самостоятельный метод оценки, а не только как вариант ММП.

## 7.1.2. Сопоставление методов

Качество методов можно определять, опираясь на различные свойства оценок. Будем делать это наиболее популярным способом, согласно которому лучше всего оценка с наименьшим средним квадратом ошибки $M[\hat{c}-c_0]^2$. Последний представляет собой сумму квадрата смещения оценки и ее дисперсии, см. также (2.16) и (2.17) в п.2.2.1.4:

$$M[\hat{c}-c_0]^2=\sigma_{\hat{c}}^2+[M\hat{c}-c_0]^2. \qquad (7.17)$$

**7.1.2.1. Пример: линейные оценки.** Рассмотрим сначала простой пример, в котором несколько разных оценок оказываются линейными по $\eta_i$ (такие оценки параметров называют *линейными*) и несмещенными. Для этого нужно, чтобы функция *f* была линейна по *c*, а помеха $\xi$ и время *t* – независимы. Этому условию удовлетворяет, например, исходный процесс

$$\eta=c_0t+\xi, \qquad (7.18)$$

где случайная величина $\xi$ распределена по нормальному закону с дисперсией $\sigma_\xi^2$. Применим методы простого усреднения, статистических моментов и МНК (совпадающий здесь с ММП) и сравним результаты.

Смысл метода простого усреднения здесь состоит в том, что через каждую точку $t_i,\eta_i$ и начало координат проводится прямая линия, определяется ее угловой коэффициент $\hat{c}_i=\eta_i/t_i$, а затем находится средний по всем полученным угловым коэффициентам (табл.7.1). Если некоторые моменты наблюдений лежат около нуля, то отклонения $\hat{c}_i$ от $c_0$ для них могут быть очень велики (т.к. они равны $\xi_i/t_i$), это может вести к очень большому разбросу значений оценки. Она может даже не иметь конечных среднего и дисперсии (табл. 7.1). Нетрудно показать, что если среднее конечно, то оценка не смещена.

Геометрический смысл метода моментов – находятся выборочные средние координаты $\langle t\rangle_N$ и $\langle\eta\rangle_N$, через полученную среднюю точку («центр тяжести») облака экспериментальных точек и начало координат проводится прямая, угловой коэффициент которой принимается за оценку $\hat{c}$. Если ее среднее конечно, то она также не смещена. Дисперсия такой оценки меньше, чем для оценки простого усреднения (табл.7.1, рис.7.3). Это отличие особенно сильно, если встречаются близкие к нулю значения координаты *t*.

Таблица 7.1.

Результаты оценки параметров в примере (7.18) различными методами. Дисперсии оценок представлены для случайной величины *t*. Для неслучайной величины *t* достаточно заменить ее «математические ожидания» на выборочные средние

| методы | формулы для $\hat{c}$ | Дисперсия $\hat{c}$ |
|---|---|---|

| | | |
|---|---|---|
| простого усреднения | $(1/N)\sum_{i=1}^{N}\eta_i/t_i$ | $\left(\sigma_\xi^2/N\right)M[1/t^2]$ |
| статистических моментов | $\sum_{i=1}^{N}\eta_i \Big/ \sum_{i=1}^{N}t_i$ | $\left(\sigma_\xi^2/N\right)M[1/\langle t\rangle_N^2]$ |
| наименьших квадратов | $\sum_{i=1}^{N}\eta_i t_i \Big/ \sum_{i=1}^{N}t_i^2$ | $\left(\sigma_\xi^2/N\right)M[1/\langle t^2\rangle_N]$ |

Что касается МНК, то его оценка имеет конечное среднее и дисперсию. Она тоже не смещена и ее дисперсия меньше, чем у оценок, полученных двумя другими методами (рис.7.3).

Для всех трех методов дисперсию помехи можно оценить как выборочную дисперсию *остатков модели* $\hat{\sigma}_\xi^2 = S(\hat{\mathbf{c}})/N$. Остатками называют величины $\varepsilon_i = \eta_i - f(t_i, \hat{\mathbf{c}})$, где $\hat{\mathbf{c}}$ – полученные оценки параметров. Заметим, что в случае нормального закона распределения $\xi$ эта оценка дисперсии помехи совпадает с ее МП-оценкой. Во всех трех случаях дисперсия $\hat{c}$ уменьшается с ростом длины ряда, т.е. оценки становятся все точнее, т.к. смещения у них в данном случае нет. Закон уменьшения дисперсии для случайных моментов наблюдений $t_i$, взятых из некоторого распределения, есть $\sigma_{\hat{c}}^2 \propto 1/N$ (рис.7.3), но при других свойствах $t_i$ он может отличаться.

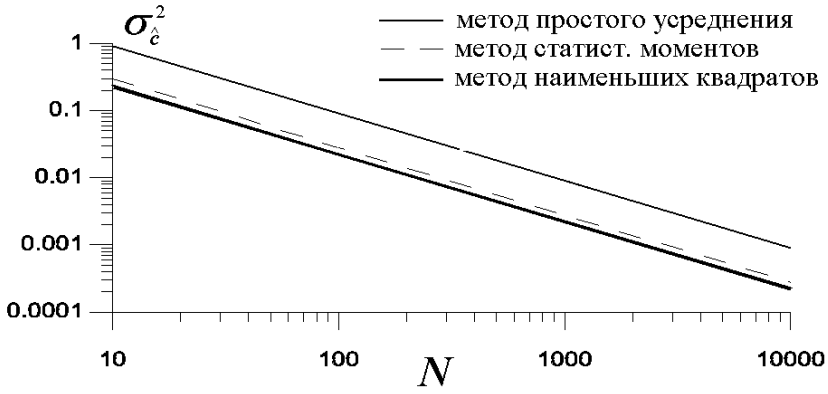



Рис.7.3. Дисперсии оценок, полученных различными методами в зависимости от длины ряда в двойном логарифмическом масштабе для примера со случайной величиной *t*, распределенной равномерно на отрезке [0.1,1.1], $\xi$ – распределена по нормальному закону с единичной дисперсией, $c_0 = 0.5$. Меньше всего дисперсия оценки МНК, затем моментной оценки, наибольшая – для простого усреднения. Закон изменения во всех случаях $\sigma_{\hat{c}}^2 \propto 1/N$, т.к. угол наклона прямых одинаков и равен 1

**7.1.2.2. Асимптотические свойства.** Рассмотрим свойства оценок при увеличении объема выборки (длины временного ряда), $N \to \infty$. Начнем со случая, когда *t* – случайная величина с конечными средним и дисперсией. При такой постановке ММП дает асимптотически несмещенные,

асимптотически эффективные, состоятельные оценки. Т.е. при достаточно большом объеме выборки это практически наилучший метод оценивания. МНК и метод моментов также дают при некоторых условиях состоятельные оценки [51, 81].

Если же моменты наблюдений $t_i$ не случайны, то ситуация меняется. Достаточно типичен случай когда $t_i$ разделены постоянным временным шагом $t_i = t_0 + i\Delta t$, и рост длины ряда $N$ соответствует увеличению длительности интервала наблюдения. Каждое новое наблюдение может тогда вносить вклад в функцию правдоподобия (7.9), отличный от предыдущих. Часто функция правдоподобия все сильнее зависит от $\eta_i$ с ростом $i$, т.е. величина $L$ в (7.9) становится *нестационарна* по $i$. Это отражается так называемым *частным информационным количеством Фишера*: в случае одного параметра это величина $I_i = (\partial f(t_i, c))/\partial c\big|_{c=c_0})^2$. Чем больше $I_i$, тем сильнее влияние наблюдения $\eta_i$ на значение получаемой МП-оценки. Такое сильное влияние имеет место при оценке параметров зависимостей

$$\eta = \sin(c_0 t) + \xi, \tag{7.19}$$

$$\eta = \sin(\exp(c_0 t)) + \xi \tag{7.20}$$

по эквидистантному ряду. В случае (7.19) МП-оценка сохраняет асимптотическую несмещенность, но ее дисперсия убывает *быстрее*: $\sigma_{\hat{c}}^2 \propto 1/N^3$, т.е. оценка становится точнее. В случае (7.20) функция правдоподобия теряет гладкость, становится слишком «изрезанной» при $N \to \infty$ (см. рис.8.4,а в п.8.1.2). Поэтому о ее асимптотических свойствах ничего не известно. При *конечных* $N$ дисперсия оценки убывает очень быстро с ростом длины ряда – приблизительно по экспоненциальному закону [287]. Но при больших $N$ поиск глобального максимума (7.9) становится практически невозможным, т.к. функция имеет слишком много локальных максимумов. Эти же проблемы встречаются и при использовании методов наименьших квадратов, наименьших модулей и минимизации наибольшего уклонения.

**7.1.2.3. Конечный объем выборки.** Учитывая сказанное выше, часто приходится ограничиваться небольшими длинами ряда, но в этом случае ММП не гарантирует оценок с наилучшими свойствами. Поэтому специальное внимание было уделено исследованию эффективности методов при работе с выборками конечного объема. При неслучайной величине $t$, нормальном распределении помехи и линейной зависимости $f$ от **c** оценки МНК (которые совпадают тогда с ММП) являются несмещенными и эффективными, т.е. наилучшими в классе несмещенных

оценок. При произвольном законе распределения помехи МНК (уже не совпадающий с ММП) дает наилучшие оценки в более узком классе – классе линейных и несмещенных оценок. Однако существуют нелинейные оценки (смещенные), которые имеют меньшую погрешность (7.17), чем оценки МНК [51].

### 7.1.3. Оптимальное оценивание

В простом примере, рассмотренном выше в п. 7.1.2.1, свойства оценок не зависели от истинных значений параметров $\mathbf{c}_0$. В общем случае такая зависимость имеет место: для одних значений $\mathbf{c}_0$ оценка может иметь малую погрешность, а для других – большую. Желательно найти метод, который дает наилучшие оценки для любого значения $\mathbf{c}_0$ из множества его возможных значений $\mathbf{C}$. Однако такого *равномерно лучшего* метода оценивания не существует. Есть оценки, оптимальные в различных смыслах, соответствующих предполагаемым способам использования метода оценивания. Два популярных варианта – метод минимакса и байесовский метод.

*Принцип минимакса.* Пусть неизвестно, при каких истинных значениях $\mathbf{c}_0$ из области $\mathbf{C}$ придется чаще использовать метод оценивания. И пусть даже в одном-единственном случае большая ошибка крайне нежелательна. Тогда нужен метод, имеющий минимальную погрешность в самом худшем случае:

$$\sup_{\mathbf{c}_0 \in \mathbf{C}} \int \left(\hat{\mathbf{c}} - \mathbf{c}_0\right)^2 p(\hat{\mathbf{c}} | \mathbf{c}_0) d\hat{\mathbf{c}} \to \min, \tag{7.21}$$

где $p(\hat{\mathbf{c}} | \mathbf{c}_0)$ – плотность распределения оценки при данном $\mathbf{c}_0$. Здесь минимизируется погрешность оценки, соответствующая наименее благоприятному для оценивания значению $\mathbf{c}_0$. Если такое $\mathbf{c}_0$ вовсе не встретится на практике, то метод, обеспечивающий (7.21), окажется слишком «осторожен» и даст оценки, не самые точные из возможных.

*Принцип Байеса.* Как мы уже обсуждали в п. 2.2.1.9, истинные значения параметров $\mathbf{c}_0$ могут рассматриваться как случайные величины.[6] Поскольку при этом одни истинные значения $\mathbf{c}_0$ встречаются чаще других, то желателен такой метод, который давал бы минимальную *среднюю* по всем $\mathbf{c}_0$ погрешность оценки, т.е.

$$\int \left(\hat{\mathbf{c}} - \mathbf{c}_0\right)^2 p(\hat{\mathbf{c}} | \mathbf{c}_0) p(\mathbf{c}_0) d\hat{\mathbf{c}} \to \min. \tag{7.22}$$

[6] В п. 2.2.1.9 рассмотрено получение оценок через апостериорную плотность, но это не единственная процедура, подпадающая под понятие байесовского подхода.

Такой метод редко дает большие ошибки (только в тех случаях, когда реализуются редкие $\mathbf{c}_0$), а часто – малые ошибки (для наиболее вероятных и, следовательно, часто встречающихся значений $\mathbf{c}_0$).

### 7.1.4. Устойчивое оценивание

Преимущество МНК и метода наименьших модулей над ММП состоит в том, что для их применения не требуется знать закон распределения помехи. Известно, что эти два метода дают оптимальные результаты, если помеха имеет распределение Гаусса или Лапласа соответственно. Возникает вопрос: сохраняются ли хорошие свойства оценок (т.е. не увеличится ли ошибка оценки) при изменениях закона распределения помехи? Если да, то какие вариации закона распределения допустимы? Если нет, как сконструировать устойчивую к этим изменениям оценку? Это предмет теории устойчивого оценивания, см. книгу [51] и ссылки в ней.

Не вдаваясь в строгие формулировки, укажем только, что оба рассматриваемых метода устойчивы в достаточно широких классах распределений. А именно, оценки наименьших квадратов устойчивы в классе законов распределения с дисперсией, не превышающей заданную конечную величину. Оценки наименьших модулей устойчивы в классе законов распределения с $p(0) \geq \Delta > 0$, где $\Delta$ – некоторое конечное число. Можно сконструировать и другие оценки, обладающие свойствами устойчивости в еще более широких классах.

### 7.1.5. Выводы

При достаточно длинном ряде, незашумленных значениях $t_i$, известном законе распределения помехи и достаточно «хорошей» (чтобы можно было найти ее глобальный максимум) функции правдоподобия, лучше всего использовать ММП. В отсутствие информации о законе распределения помехи и линейной зависимости $f$ от параметров МНК и метод наименьших модулей – наилучшие каждый в своем классе. Но МНК реализовать проще, т.к. задача оптимизации линейна, в отличие от прочих методов. МНК часто применяется и при нелинейной зависимости $f$ от параметров, хотя тогда для его реализации нужно решать более трудную задачу нелинейной оптимизации.

Метод статистических моментов в целом уступает МНК по эффективности. Еще хуже, как правило, метод простого усреднения. Но есть ситуации, когда они могут иметь свои преимущества [287].

Мы только кратко коснулись вопросов устойчивого и оптимального оценивания, поскольку соответствующие задачи более сложны и пока практически не рассматриваются при моделировании по временным рядам.

## 7.2. Аппроксимация

### 7.2.1. Две постановки задачи и термины

Аналогично п. 7.1 рассмотрим две ситуации – детерминированную и стохастическую. В «детерминированном» случае исходный процесс есть

$$\eta = F(t). \tag{7.23}$$

Вид «истинной» функции $F$ неизвестен, и требуется найти модельную функцию $f$, как можно точнее приближающую $F(t)$ на интересующей нас области значений $t$. Задача приближения одной функции $F(t)$ другой функцией $f(t)$ на некотором отрезке $[a, b]$ – центральная в теории аппроксимации[7] [22]. В широком смысле под *аппроксимацией* понимают *замену одного объекта другим, в некотором смысле близким к исходному*.[8] Если построенная по данным наблюдений модельная функция $f$ используется для расчета значений величины $\eta$ (т.е. для замены $F$) в промежуточные моменты времени $t_1 < t < t_N$, то говорят об *интерполяции* зависимости $\eta(t)$. Если же аппроксимирующая функция используется для расчета значений $\eta$ за пределами наблюдаемого интервала ($t < t_1$ или $t > t_N$), то говорят об *экстраполяции* зависимости $\eta(t)$.[9] В частности,

---

[7] Синонимы – теория приближения функций, конструктивная теория функций.

[8] Пример: в астрономии Птолемея движения небесных тел относительно Земли аппроксимировались с помощью комбинации движений по окружностям – эпициклам, центры которых движутся по другим окружностям – деферентам, центр которых – Земля. Для достижения высокой точности прогнозов будущих положений планет на небесной сфере, что было важно в морской навигации для контроля курса корабля, была развита очень сложная система эпициклов разного радиуса. Эта работа стимулировала развитие сферической геометрии в средние века. Когда на смену геоцентрической системе пришла гелиоцентрическая система Коперника, то сначала последняя значительно уступала своей предшественнице в отношении точности прогнозов и на практике была менее эффективна. Но с дальнейшим развитием взглядов и моделей гелиоцентрической системы (особенно с появлением закона всемирного тяготения Ньютона) движение светил стало уже значительно лучше описываться в новой теории, и она вытеснила старую.

[9] В Математическом энциклопедическом словаре дается более широкое определение интерполяции – *это приближенное или точное нахождение какой-либо величины по известным отдельным значениям этой же величины или других величин, связанных с ней* [118]. При таком определении интерполяция включает в себя и аппроксимацию. В случае экстраполяции в классической математике (чаще в комплексном анализе) используется, напротив, несколько более узкое толкование: *это продолжение функции за пределы ее области определения, при котором продолженная функция (как правило, аналитическая) принадлежит заданному классу*. Для нас удобнее использовать приведенные в тексте и принятые среди специалистов по численным методам определения.

экстраполяцию на «область будущего» $t > t_N$ называют прогнозом в узком смысле.

В «стохастическом» случае исходный процесс описывается как

$$\eta = F(t) + \xi(t), \tag{7.24}$$

где вид функции $F$ неизвестен, а $\xi$ – случайная величина с нулевым средним, не зависящая от $t$. Для характеристики таких процессов используют некоторые дополнительные понятия, отличные от детерминированного случая. Полная информация об исходном процессе содержится в функции условной плотности распределения вероятностей $p_0(\eta|t)$ при фиксированном $t$. Но «истинная» функция $p_0(\eta|t)$ неизвестна. Для модельного описания можно стремиться подобрать закон $p(\eta|t)$, как можно точнее приближающий $p_0(\eta|t)$. Но задача восстановления всего закона распределения по конечной выборке трудна, и такая полная информация не всегда нужна. Часто вполне достаточно знать условное среднее $M[\eta|t]$ величины $\eta$ при данном $t$ и разброс значений $\eta$ относительно $M[\eta|t]$, характеризуемый условной дисперсией. Зависимость условного среднего величины $\eta$ от $t$ есть детерминированная функция, которая называется *регрессией*.[10]

Регрессия наиболее точно (с минимальным средним квадратом ошибки) предсказывает $\eta$ по $t$: $\varepsilon^2(F) = \int (\eta - F(t))^2 p(\eta|t) d\eta = \min\limits_f \varepsilon^2(f)$. В случае (7.24) имеем $M[\eta|t] = F(t)$, т.к. $M[\xi] = 0$, т.е. $F(t)$ является регрессией. При моделировании обычно ставится задача подобрать функцию $f(t)$, как можно точнее приближающую «истинную» регрессию $F(t)$. Эта задача носит названия «восстановления регрессии» или «оценки регрессии».

Итак, в обеих постановках возникает задача аппроксимации (восстановления, оценки) функции $F(t)$ по наблюдаемым данным. Как правило, модель ищется в некотором классе $f(t, \mathbf{c})$ и нужно найти $P$-мерный вектор параметров $\mathbf{c} = \hat{\mathbf{c}}$, который обеспечивает максимальную

[10] Этот термин был впервые использован английским статистиком Ф. Гальтоном (1866) в теории наследственности. Он изучал, как зависит рост детей $Y$ от роста родителей $X$, и установил, что если рост родителей превышает среднее значение на $b$ единиц, то рост детей превышает это среднее значение меньше, чем на $b$ единиц. Это явление он назвал регрессией, что означает в переводе с латыни «обратное движение». Найденная им зависимость условного среднего $Y$ от $X$ тоже была названа регрессией

близость модельной функции $f$ и истинной $F$ (п.7.2.2).[11] При этом очень важно выбрать оптимальный размер модели (п.7.2.3) и подходящий класс аппроксимирующих функций (п.7.2.4). Иногда это можно сделать, опираясь на визуальное рассмотрение наблюдаемого сигнала и подбирая элементарные функции с похожими графиками. Но в общем случае используется аппроксимация в некотором функциональном базисе.

### 7.2.2. Расчет параметров

Параметры модели **c** нужно подобрать так, чтобы наилучшим образом обеспечить условие $f(t,\hat{\mathbf{c}}) \approx F(t)$. Причем близость функций желательна не только в моменты наблюдений $t_1,...t_N$, но и в *промежуточные* моменты, а иногда даже в прошлые или будущие (последнее, однако, весьма затруднительно). Для количественной характеристики близости вводится соответствующая мера расстояния – метрика $\rho$ в пространстве функций[12]. Наиболее популярный подход – использование в качестве $\rho$ среднего квадрата отклонения с весовой функцией $p(t)$:

$$\rho(f,F) = \int \left(f(t,\mathbf{c}) - F(t)\right)^2 p(t)dt\,. \qquad (7.25)$$

Физический смысл такой метрики – средний по $t$ квадрат отклонения модельной функции $f(t,\mathbf{c})$ от $F(t)$, если $t$ – случайная величина с плотностью распределения $p(t)$. Для случая (7.24) расстояние $\rho$ с точностью до дисперсии помехи равно среднему квадрату отклонения $f(t,\mathbf{c})$ от $\eta(t)$:

---

[11] Даже если истинная регрессия $F(t)$ попадает в выбранный класс $f(t,\mathbf{c})$, т.е. $F(t) = f(t,\mathbf{c}_0)$ для некоторого $\mathbf{c}_0$, то задачу восстановления $F(t)$ нельзя «подменить» задачей нахождения как можно более точной оценки параметров $\mathbf{c}_0$. Дело в том, что наилучшая оценка параметров в «истинном» классе $f(t,\mathbf{c})$ не обязательно обеспечивает наилучшее качество приближения $F$, поскольку наилучшая аппроксимация может быть достигнута в другом классе [51].

[12] К вопросу о корректности постановки задачи аппроксимации. Предполагается, что *существуют* значения параметров $\hat{\mathbf{c}}$, которые доставляют минимум расстоянию $\rho(f,F)$. Для полного комфорта желательно, чтобы функции вида $f(t,\mathbf{c})$ были *плотны* в классе функций, к которому принадлежит $F(t)$. Это означает, что в окрестности почти каждой $F(t)$ из этого класса найдутся нужные нам $f(t,\mathbf{c})$, с помощью которых ее удастся аппроксимировать с заданной точностью. Наконец, если $f(t,\mathbf{c})$ образуют чебышевское множество [112], т.е. содержащее все предельные точки последовательностей функций $f(t,\mathbf{c})$, оптимальная аппроксимация будет и единственной! Примером служат радиальные базисные функции (см. п.10.2.1.2).

$$\iint\left(f(t,\mathbf{c})-\eta(t)\right)^2 p(t)\,p(\eta|t)dtd\eta=\iint\left(f(t,\mathbf{c})-F(t)-\xi\right)^2 p(t)\,p(\xi)dtd\xi=$$
$$=\rho(f(t,\mathbf{c}),F(t))+2\left(\int\left(f(t,\mathbf{c})-F(t)\right)p(t)dt\right)\left(\int\xi p(\xi)d\xi\right)+\int\xi^2 p(\xi)d\xi= \qquad (7.26)$$
$$=\rho(f(t,\mathbf{c}),F(t))+\sigma_\xi^2.$$

Возможны и другие подходы к заданию $\rho$, возможна не случайная величина $t$. Но рассуждения во всех этих случаях аналогичны с некоторыми оговорками, так что мы ограничимся только вариантом (7.25).

Итак, требуется найти значения параметров, которые минимизируют функционал (7.25). Поскольку «истинная» плотность вероятностей $p(t)$ в (7.25), как правило, неизвестна, необходимо заменить минимизацию (7.25) задачей минимизации функционала, который можно рассчитать по наблюдаемым данным и точка минимума которого близка к точке минимума (7.25). Аналогично (7.12), таким *эмпирическим* функционалом может служить величина среднего квадрата ошибки прогноза на наблюдаемом ряде:

$$\varepsilon^2(\mathbf{c})=S(\mathbf{c})/N=\frac{1}{N}\sum_{i=1}^{N}\left(\eta_i-f(t_i,\mathbf{c})\right)^2. \qquad (7.27)$$

Параметры модели $\hat{\mathbf{c}}$ определяются из условия минимума (7.27). Это вариант так называемого *метода минимизации эмпирического риска* [51]. В данном случае $\varepsilon^2(\mathbf{c})$ имеет смысл суммы дисперсии шума и квадрата ошибки аппроксимации (7.26). Если класс функций $f(t,\mathbf{c})$ удачно выбран, так что в нем есть функции, очень близкие к $F$ в смысле метрики $\rho$, то величина $\hat{\varepsilon}^2=\varepsilon^2(\hat{\mathbf{c}})$ есть почти не смещенная оценка дисперсии помехи $\xi$.

Как видно из (7.27), метод расчета параметров здесь – это обычный МНК, т.е. в техническом плане отличий между задачами пп. 7.1 и 7.2 нет.

### 7.2.3. Выбор размера модели, переобучение и «бритва Оккама»

Не касаясь пока выбора вида функции $f$ (п. 7.2.4), проиллюстрируем проблему выбора размера модели в рамках заданного класса функций на классическом примере аппроксимации алгебраическим многочленом:

$$f(t,\mathbf{c})=c_1+c_2t...+c_{K+1}t^K. \qquad (7.28)$$

Размер модели определяется здесь числом свободных[13] параметров многочлена. В простейшем варианте, когда все параметры (7.28) свободны, приходим к известной в статистике (в полиномиальной регрессии) задаче

[13] Свободными называют параметры, оцениваемые по временному ряду, на которые не наложено дополнительных ограничений – равенств.

выбора порядка многочлена $K = P + 1$. Теоретическое обоснование для использования алгебраического многочлена дает знаменитая теорема Вейерштрасса, согласно которой любая непрерывная на отрезке функция может быть сколь угодно точно равномерно приближена алгебраическим многочленом. Отсюда следует, что с помощью многочлена достаточно высокого порядка $K$ можно обеспечить сколь угодно малую величину $\rho(f, F)$.

**7.2.3.1. Понятия недообученной и переобученной модели.** Какую именно величину $K$ выбрать на практике? Очень малый порядок многочлена зачастую не годится, т.к. не дает возможности с хорошей точностью аппроксимировать сложные зависимости. В таких случаях говорят, что модель *недообучена*. Но, как мы покажем ниже, и слишком высокие порядки не годятся. Важно строить достаточно «экономичную» модель!

Проверенный способ действия состоит в том, чтобы строить модели с многочленами $f$ различных порядков, начиная с 0, и остановиться на том значении $K$, когда дальнейшее увеличение перестает приводить к улучшению модели по некоторому критерию. Таких критериев имеется несколько: минимум тестовой ошибки $\hat{\varepsilon}^2_{test} = \varepsilon^2_{test}(\hat{\mathbf{c}})$; насыщение эмпирической ошибки $\hat{\varepsilon}^2$; «скользящий контроль»; минимум некоторой целевой функции, которая является суммой эмпирической ошибки $\hat{\varepsilon}^2$ и штрафного слагаемого (см. п. 7.2.3.5).

Все эти критерии направлены на «борьбу с некорректностью» задачи, проявляющейся в том, что существует бесконечно много модельных функций, способных одинаково хорошо описать *конечный* набор экспериментальных точек на плоскости $(t, \eta)$. Именно по этой причине нельзя использовать для выбора порядка многочлена $K$ простой и, казалось бы, логичный критерий минимума эмпирической ошибки $\hat{\varepsilon}^2$. Дело в том, что $\hat{\varepsilon}^2$ *всегда* является невозрастающей функцией порядка $K$. Она равна нулю, когда число коэффициентов многочлена $P$ равно числу точек ряда $N$ (и все $t_i$ различны), т.е. график многочлена проходит *в точности* через экспериментальные точки $(t_i, \eta_i)$. Но такая модель, как правило, крайне плоха. Она «научилась» точно воспроизводить только наблюдаемую случайную реализацию вместе со всеми ее шумами (рис.7.4,а). Из-за этого она будет очень плохо предсказывать в среднем новые наблюдения, т.к. при новых наблюдениях конкретные значения помехи будут другими. Полученная таким образом функция $f(t, \hat{\mathbf{c}})$ как оценка $F(t)$ обладает огромной дисперсией. Даже если эта оценка несмещена, ее случайная погрешность может быть сколь угодно велика. Говорят, что такая модель обладает плохой способностью к *обобщению* информации, она

*переобучена*. Это основное проявление некорректности постановки задачи. Избежать переобучения модели – одна из важнейших проблем при эмпирическом моделировании!

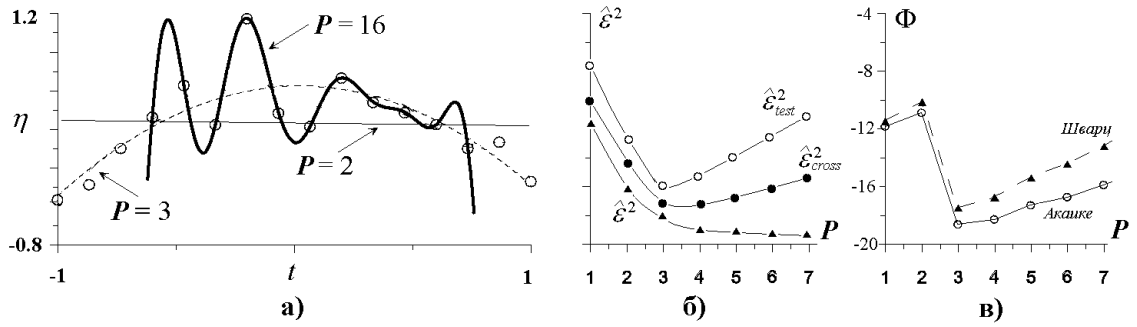



Рис.7.4. Аппроксимация зависимости при квадратичной функции $F(t)$ по ряду длиной $N = 16$ точек многочленами различных порядков. а) Графики модельных функций различных порядков $K = P - 1$. Тонкая линия – недообученная модель, жирная линия – переобученная. Штриховая линия – оптимальная. б) Выбор оптимального порядка по ошибкам прогноза (общая иллюстрация). в) Выбор оптимального порядка по критериям Шварца и Акаике. Результаты расчетов для примера, показанного на рисунке (а)

**7.2.3.2. Критерий минимума тестовой ошибки.** Если в распоряжении исследователя есть еще один временной ряд того же процесса $t'_i, \eta'_i, i = 1,...,N'$ (тестовый), то хорошим критерием выбора порядка многочлена является минимум тестовой ошибки аппроксимации для модели с заданным порядком многочлена, полученной по тренировочному ряду,

$$\hat{\varepsilon}^2_{test} = \frac{1}{N'} \sum_{i=1}^{N'} \left(\eta'_i - f(t'_i, \hat{\mathbf{c}})\right)^2 . \qquad (7.29)$$

Причем можно пользоваться еще и следующим соображением: для непереобученной модели величины $\hat{\varepsilon}^2$ и $\hat{\varepsilon}^2_{test}$ примерно равны. Минимизация (7.29) – самый надежный способ выбора $K$ (рис.7.4,б, светлые кружки), поскольку он основан на расчете не по той выборке, по которой строилась модель. Такие критерии называются *вневыборочными*.

**7.2.3.3. Критерий насыщения «тренировочной» ошибки.** Однако тестовый ряд не всегда есть в наличии, и приходится извлекать необходимую информацию из имеющегося тренировочного ряда. Соответствующие критерии называют *внутривыборочными*. Если по характеру уменьшения $\hat{\varepsilon}^2$ с ростом $K$ видно, что при значениях $K$, больших некоторого «порогового», погрешность меняется слабо – наблюдается насыщение, то это пороговое значение и можно выбрать как наилучшее (рис.7.4,б, треугольники). Нередко такой подход может дать

завышенное значение порядка, по сравнению с действительно оптимальным.

**7.2.3.4. Метод «скользящий контроль».** Другое название метода – метод «перекрестной проверки». Он занимает промежуточное положение между описанными выше. Идея его следующая [51]. Из имеющегося временного ряда поочередно исключается каждое отдельное наблюдение $t_i, \eta_i$. По оставшемуся ряду длиной $N-1$ строится модель, обозначим ее $f(t, \hat{\mathbf{c}}_i)$. Рассчитывается ошибка прогноза исключенного наблюдения с помощью полученной модели $\hat{\varepsilon}_i = \eta_i - f(t_i, \hat{\mathbf{c}}_i)$. Наконец, рассчитывается среднеквадратичная ошибка «перекрестного» прогноза $\hat{\varepsilon}^2_{cross} = \frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_i^2$. Оптимальный порядок определяется из условия минимума $\hat{\varepsilon}^2_{cross}$ (рис.7.4,б, черные кружки). Этот подход также может дать завышенное значение *K*, но более надежен, чем критерий насыщения $\hat{\varepsilon}^2$.

**7.2.3.5. Критерии минимума целевых функций со штрафным слагаемым.** Ряд методов автоматического выбора порядка многочлена опирается на идею минимизации целевой функции вида

$$\Phi(P) = g_1(\hat{\varepsilon}^2) + g_2(P), \tag{7.30}$$

где $g_1, g_2$ – возрастающие функции своих аргументов. Первое слагаемое определяет вклад эмпирической ошибки, а второе – размера модели. Эти методы развиты в теории информации и теории статистического оценивания из разных соображений, часто привлекающих принцип максимума правдоподобия. Наличие минимума (7.30) следует ожидать зачастую при промежуточных значениях размера модели, т.к. при малом порядке многочлена слишком велика эмпирическая ошибка, а при большом – слишком велико становится второе слагаемое.

Целевая функция $\Phi(P) = (N/2)\ln\hat{\varepsilon}^2 + P$ называется критерием Акаике [184], $\Phi(P) = (N/2)\ln\hat{\varepsilon}^2 + P(\ln N/2)$ – критерием Шварца [304], $\Phi(P) = \ln\hat{\varepsilon}^2 + P$ – энтропией модели [216]. Более громоздка формула для такой целевой функции, как *длина описания* [250, 293]. Минимизация длины описания – самый популярный в настоящее время подход, полученный из соображений оптимального сжатия информации. Он имеет прямое отношение к понятию *алгоритмической сложности* процесса. Кроме того, есть метод упорядоченной минимизации риска, где «штрафное» слагаемое зависит от «емкости» класса аппроксимирующих функций (характеристики его «широты») [51]. На рис.7.4,в проиллюстрированы два подхода. Для конкретного случая все методы

(7.30) могут дать несколько различные ответы на вопрос об оптимальном размере модели. Поэтому их рекомендации следует воспринимать критично, только как грубую оценку, и проверять качество моделей, построенных при нескольких близких значениях $K$.

**7.2.3.6. Бритва «Оккама».** Все подходы к выбору наилучшего размера модели (минимально возможного для удовлетворительного описания данных) реализуют *принцип экономичности* при моделировании [45, 250]. Это конкретный вариант общего научно-философского принципа: «*Не следует умножать число сущностей без необходимости*». Этот принцип носит название *бритвы Оккама*, [14] хотя похожие утверждения высказывались уже Аристотелем, см., например, [305].

**7.2.3.7. Оптимизация структуры модели.** При любом из вышеописанных подходов итоговая модель с многочленом порядка $K$ обязательно содержит все слагаемые до порядка $K$ включительно. Но нетрудно представить ситуацию, когда некоторые из промежуточных степеней лишние и их лучше бы исключить из модели. Этот недостаток возникает из-за того, что слагаемые добавляются в модель в наперед заданном порядке.

В общем случае требуется разрабатывать методы поиска оптимальной модели размера $P$, где рассматриваются всевозможные комбинации $P$ функций-слагаемых из более широкого множества $P_{\max}$ функций. Это очень просто, если система базисных функций ортогональна на множестве наблюдений. Но может быть очень сложно в противном случае, т.к. прямой перебор всех комбинаций невозможен из-за *комбинаторного взрыва* (астрономически большое число возможных комбинаций даже для умеренных значений $P_{\max}$ и $P$). Существует множество методов [278], некоторые из них мы только упомянем, а один из них подробнее опишем в главе 8: последовательное добавление в модель базисных функций по скорейшему снижению погрешности аппроксимации [250]; исключение из изначально «большой» модели так называемых «лишних слагаемых» по самому медленному увеличению погрешности аппроксимации [183], по критерию Стьюдента [87], по большой вариации коэффициентов при использовании различных тренировочных рядов [25, 190]. Поиск оптимальной комбинации функций-слагаемых называют *оптимизацией структуры* модели.

---

[14] *Pluralitas non est ponenda sine necessitate*. Уильям Оккам (1285-1349) – знаменитый английский философ и логик.

## 7.2.4. Выбор класса аппроксимирующих функций

**7.2.4.1. Примеры функциональных базисов.** При моделировании могут использоваться любые виды функций, но гораздо проще выполнять расчеты, когда параметры входят в модель линейно, т.е.

$$f(t,\mathbf{c}) = c_1\phi_1(t) + \ldots + c_P\phi_P(t), \tag{7.31}$$

где $\phi_k$, называются *базисными функциями*. Обычно это – элементы бесконечной системы функций $\phi_k, k = 1,2,\ldots$, в которой можно сколь угодно точно приблизить любую «достаточно хорошую» (например, непрерывную) функцию *F*. Другими словами, эта система есть функциональный базис в пространстве функций с определенными свойствами. Функцию (7.31) называют *обобщенным многочленом* по системе функций $\phi_1(t),\ldots,\phi_P(t)$, а также *псевдолинейной* моделью (приставка «псевдо» подчеркивает, что линейна зависимость от параметров, а не от аргумента). В примере (7.28) для аппроксимации использовался стандартный полиномиальный базис, а базисными функциями были одночлены $t^{k-1}$.

Известна также тригонометрическая система базисных функций $1, \cos\omega t, \sin\omega t, \cos 2\omega t, \sin 2\omega t, \ldots$. В случае равномерной выборки решение задачи на наименьшие квадраты достигается с помощью прямого преобразования Фурье (п. 6.4.2.1). Значения коэффициентов $c_k$ приобретают физический смысл: через них выражаются компоненты спектра мощности сигнала. Использование тригонометрического многочлена для аппроксимации позволяет избежать больших выбросов, свойственных алгебраическому многочлену высокого порядка. Оно максимально эффективно, когда в анализируемом сигнале ярко проявляется «повторяемость» (он периодичен или почти периодичен), рис.7.5,б. Алгебраические же многочлены лучше использовать для плавно меняющихся сигналов (пусть и замысловатой формы) без периодичностей, импульсов и резких изменений, рис.7.5,а.

Еще один популярный вид псевдолинейных моделей – вейвлеты (базисные функции – «импульсы» разной ширины и в разных положениях на временной оси, см. п. 6.4.2.2). Этот класс функций дает гибкий инструмент для описания сигналов с меняющимся во времени частотным составом, сигналов импульсного характера (рис.7.5,в) и т.д.

Используются и комбинации функциональных базисов [72].

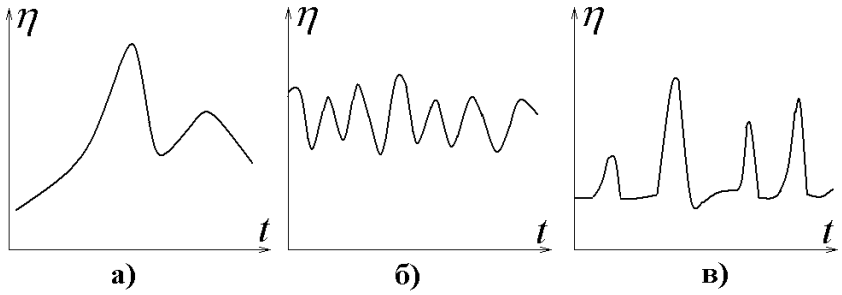



Рис.7.5. Примеры сигналов, для аппроксимации которых лучше использовать алгебраический многочлен (а), тригонометрический многочлен (б), вейвлет (в)

**7.2.4.2. Нелинейная зависимость от параметров.** Используются и аппроксимирующие функции, нелинейно зависящие от параметров: радиальные, цилиндрические и эллиптические базисные функции [308] (п. 10.2.1.2), искусственные нейронные сети[15] [112] (пп.3.8, 10.2.1.3), специально подобранные функции. При нелинейной зависимости от параметров задача минимизации (7.27) не сводится к линейной системе уравнений. Ее приходится решать, используя численные итерационные методы минимизации. Широко применяется гладкая оптимизация [73]: градиентный спуск, методы Ньютона, Гаусса – Ньютона, квазиньютоновские методы. Но целевая функция может иметь много минимумов, и только один из них (глобальный) дает наилучшие значения параметров. В зависимости от начального приближения можно попасть в тот или иной локальный минимум. Не существует универсальных способов избежать этого. Проблема локальных минимумов становится тем сложнее, чем больше число неизвестных параметров. Решают ее (без гарантий успеха) перебором различных стартовых догадок или с помощью более сложных подходов стохастической оптимизации (генетические методы, метод имитации отжига, и др.) [79, 216].

**7.2.4.3. Понятие о глобальных и локальных методах.** Если аппроксимирующая функция задается с помощью единой формулы во всем диапазоне изменений аргумента, как в случае многочлена, то аппроксимацию (и модель) называют *глобальной* [206]. К ним относятся все вышеупомянутые структуры. Альтернативным и часто не менее эффективным является *локальный* (кусочный) подход, когда аппроксимирующая функция задается некоторой (чаще простой) формулой со своим набором параметров для каждой небольшой области значений аргумента [206, 224]. Наиболее популярные примеры – кусочно-постоянные и кусочно-линейные аппроксимации и кубические сплайны [85, 153]. Локальные модели лучше описывают менее гладкие зависимости

[15] Заметим, что сигмоидальные функции, которые обычно используются в ИНС, не образуют чебышевского множества [112].

с изломами и разрывами. Но они более чувствительны к действию шума, чем глобальные с небольшим числом параметров. Поэтому локальные модели лучше при большом объеме данных и низком уровне шумов для сколь угодно сложной нелинейности, а глобальные модели имеют преимущества при меньшем объеме данных и более высоком уровне шума для систем с не очень сложной нелинейностью.

**7.2.4.4. Какой класс модельных функций «лучше»?** Теоретически, можно использовать любой функциональный базис для аппроксимации любой достаточно хорошей функции $F(t)$. Практически, разные базисы имеют свои преимущества и недостатки. Для аппроксимации конкретной зависимости $F(t)$ в разных базисах имеется разный оптимальный размер модели. Наилучший базис для аппроксимации $F(t)$ – тот, в котором оптимальный размер модели наименьший, т.е. достаточно небольшого числа базисных функций. При этом снижается опасность переобучения модели и уменьшаются ошибки оценок ее параметров. При неудачном выборе базиса может не хватить данных, чтобы надежно оценить параметры модели с необходимым большим числом базисных функций.

Если исследователь, опираясь на собственный опыт и априорную информацию об объекте, выбирает структуру модели в специальном виде, то в случае успеха дополнительным преимуществом может оказаться наличие физического смысла ее параметров. Но изредка такой смысл может быть и в универсальных конструкциях. Отметим энергетическую трактовку амплитуд гармоник в тригонометрическом многочлене и обычно отсутствующий физический смысл коэффициентов алгебраического многочлена.

## 7.3. Диагностическая проверка модели

В рассмотренных выше постановках задачи проверка адекватности основывается на изучении *остатков модели*, т.е. величин $\varepsilon_i = \eta_i - f(t_i, \hat{\mathbf{c}})$. Их называют также *невязками*. Желательно проверять свойства остатков на тестовом ряде, но чаще приходится ограничиваться тренировочным. Для постановки (7.1) – модель с известной структурой, нет шума – все остатки должны быть равны нулю (с машинной точностью).

При наличии шумов – (7.3) или (7.24) – проверяются статистические свойства остатков. Обычно, уже при выборе метода оценки параметров делают предположения о свойствах шума $\xi$. Остатки модели должны удовлетворять этим предположениям, чтобы подтвердить правомерность использования метода и надежность результатов. Чаще всего предполагают *статистическую независимость* значений помехи и ее *нормальность*.

### 7.3.1. Проверка независимости остатков

Есть различные тесты на независимость: критерий восходящих и нисходящих серий; критерий серий, основанный на медиане выборки; критерий квадратов последовательных отношений [1]. Они пригодны при любом законе распределения $\xi$. Мы остановимся подробнее на расчете автокорреляционной функции (АКФ), чтобы проверить *некоррелированность* остатков (строго говоря, подход применим, только если выборка равномерна). Это более слабое свойство, но в случае нормального распределения свойства статистической независимости и некоррелированности совпадают.

Оценка АКФ остатков: $\hat{\rho}(n) = \sum_{i=1}^{N-n} \varepsilon_i \varepsilon_{i+n} \Big/ \sum_{i=1}^{N} \varepsilon_i^2$, $n = 0,1,...,K < N$. Обычно принимают $K \le N/4$. Для последовательности независимых величин теоретические значения: $\rho_\xi(0) = 1, \rho_\xi(n) = 0, n > 0$. Рассмотрим график $\hat{\rho}(n)$ (рис.7.6,а,г) и оценим его отклонение от теоретических значений. В предположении о независимости остатков, значения $\hat{\rho}_\xi(n), n > 0$ распределены почти по нормальному закону (строже, по закону $\chi^2$ с числом степеней свободы $N - n$) и лежат в интервале $\pm 2\hat{\sigma}_\xi / N$ (показан на рис.7.6,а,г пунктиром) с вероятностью 0.95 [45]. Независимость остатков не опровергается, т.к. не менее 95% значений лежат в пределах интервала.

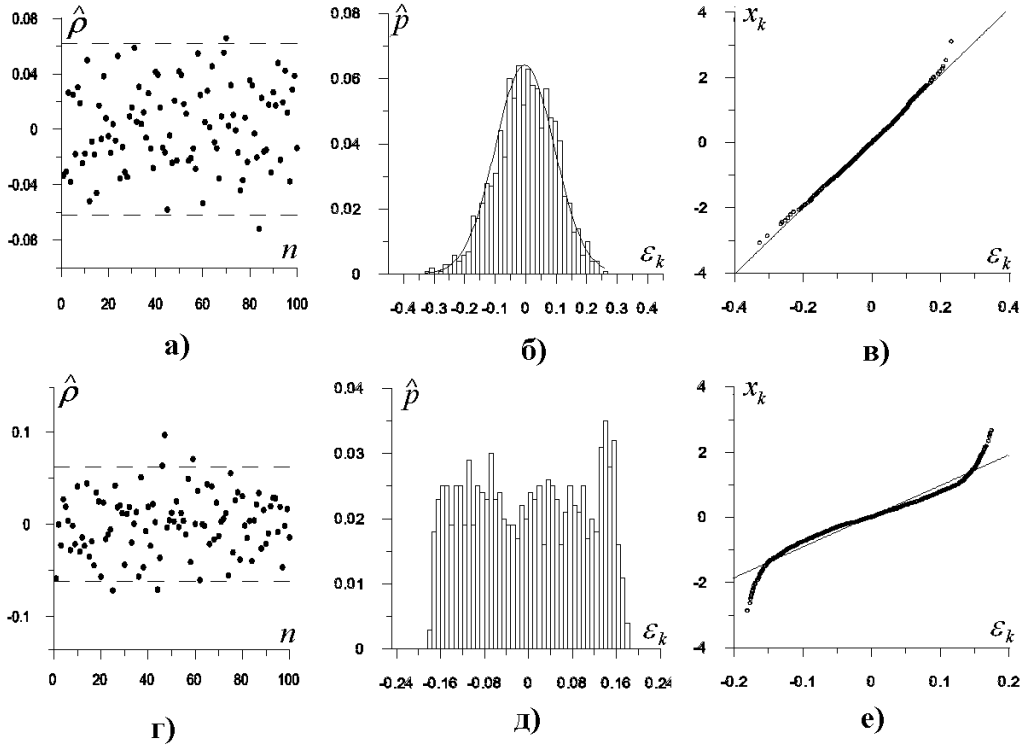
### 7.3.2. Проверка нормальности остатков

Для оценки нормальности распределения остатков можно использовать количественные статистики: тест Колмогорова – Смирнова, критерий $\chi^2$ и т.д. [1, 289]. Мы ограничимся двумя графическими способами визуальной проверки (рис.7.6).

Самый простой способ – построить гистограмму остатков (т.е. оценку плотности распределения). Для этого весь диапазон изменения $\hat{\xi}_i$ разбивается на $M$ интервалов (бинов), обычно одинаковых по размеру. Подсчитывается частота попадания величины в каждый интервал. Над каждым интервалом строится прямоугольник с основанием, равным ширине интервала, и с высотой, равной отношению частоты попадания в этот интервал к ширине интервала. Визуально можно оценить «похожесть» гистограммы на нормальное распределение, см. два примера на рис.7.6,б,д.

Однако проверить эту «похожесть» визуально трудно, тем более, что гистограмма как оценка плотности распределения имеет большую дисперсию при малой ширине бинов: задача восстановления плотности вероятности по выборке некорректна [51]. Более удобным является

визуализация графика на *нормальной вероятностной бумаге* [45]. Для этого по оси абсцисс откладываются наблюдаемые величины остатков $\varepsilon_i$, а по оси ординат – соответствующие значения $x_i$, рассчитываемые по функции распределения остатков (восстановление функции распределения – задача корректная [51]). Если остатки $\varepsilon_i$ распределены нормально с нулевым средним, то точки на графике лягут на прямую, которая проходит через начало координат, а угол наклона определяется дисперсией $\varepsilon_i$.[16] Оценить, ложатся ли точки на прямую, гораздо легче, чем оценить схожесть гистограммы с нормальным законом (рис.7.6,в,е). Представленные критерии независимости и нормальности проиллюстрированы на практических примерах в [39].




[16] Каждому из значений $\varepsilon_k$ соответствует свое значение эмпирической функции распределения остатков $\hat{\Phi}(\varepsilon_k)$. Если все $\varepsilon_i$ различны, $\hat{\Phi}(\varepsilon_k)$ можно оценить как отношение числа значений $\varepsilon_i$, меньших или равных $\varepsilon_k$, к их общему числу *N*. Обозначим $\Phi_0(x)$ – функцию распределения стандартного нормального закона (нулевое математическое ожидание и единичная дисперсия). Выберем такое значение $x_k$, при котором $\Phi_0(x_k) = \hat{\Phi}(\varepsilon_k)$. Это значение единственно, т.к. $\Phi_0(x)$ – непрерывная строго монотонная функция. Поставим в соответствие $\varepsilon_k$ величину $x_k$, т.е. $x_k = \Phi_0^{-1}(\hat{\Phi}(\varepsilon_k))$. График зависимости $x_i(\varepsilon_i)$ и есть график на нормальной вероятностной бумаге.

Рис.7.6. Анализ остатков, полученных при построении модельного многочлена первого порядка с помощью МНК по ряду длиной $N$ = 1000 точек от процесса $\eta(t) = 1 + t + \xi$ (7.26) с нормальным (а-в) и равномерно распределенным (г-е) белым шумом с дисперсией 0.01. Левый столбец – оценки АКФ (некоррелированность не опровергается, т.к. значения корреляций выходят за пределы 95%-го доверительного интервала не чаще, чем в 5% случаев). Средний столбец – гистограммы. В соответствии со свойствами шума в исходном процессе, только в первом случае гистограмма похожа на плотность распределения нормального закона. Правый столбец – графики на нормальной вероятностной бумаге. Во втором случае нормальность остатков опровергается

При наличии тестового ряда признаком неадекватности модели может служить различие выборочных дисперсий остатков на тренировочном и тестовом рядах. Вывод о неадекватности модели может быть связан с ее переобучением или недообучением, с неудачным выбором класса аппроксимирующих функций, с нахождением локального, а не глобального минимума целевой функции, или с тем, что вероятностные свойства шумов в исходном процессе не были «угаданы» и был выбран не лучший метод оценки параметров. Наконец, может быть неадекватна структура модели, например, шумы в (7.24) являются на самом деле зависимыми друг от друга и т.п. Следует выяснить, какая из этих причин имеет место, и изменить действия на соответствующем этапе процедуры моделирования (рис.5.1).

## 7.4. Примеры применения моделей

Несмотря на простоту рассмотренных в данной главе моделей, умение строить их важно в практике моделирования. Они имеют и самостоятельное значение, а чаще выступают как элементы при решении более сложных задач. Рассмотрим два приложения: прогноз (одна из центральных тем книги) и численное дифференцирование (важно при построении модельных дифференциальных уравнений, см. лабораторные работы [30, 37]).

### 7.4.1. Прогноз

Необходимо предсказать значение $\eta$ в заданный момент времени $t$. Причем важно не только дать «точечный» прогноз $\hat{\eta}(t)$, но и получить оценку его погрешности, т.е. указать интервал, в котором с высокой вероятностью должно лежать значение $\eta(t)$. Это – «интервальный» прогноз. Часто интервал ищут в виде $\hat{\eta}(t) \pm \Delta\hat{\eta}(t)$. Для простоты рассмотрим задачу в случае единственного оцениваемого параметра.

Чаще всего в качестве точечного прогноза $\hat{\eta}(t)$ для всех рассмотренных в данной главе моделей принимают величину $\hat{\eta}(t) = f(t, \hat{c})$. Ошибка такого прогноза $e(t)$ равна

$$e(t) \equiv \hat{\eta}(t) - \eta(t) = [f(t, \hat{c}) - f(t, c_0)] - \xi(t). \qquad (7.32)$$

При малой погрешности оценки параметра или в случае псевдолинейной модели это выражение можно переписать в виде

$$e(t) = k(t) \cdot (\hat{c} - c_0) - \xi(t), \qquad (7.33)$$

где $k(t) = \partial f(t, c) / \partial c \big|_{c=c_0}$. Погрешность прогноза удобно характеризовать средним квадратом (7.33):

$$M[e^2(t)] = k^2(t) \cdot M[\hat{c} - c_0]^2 + \sigma_\xi^2. \qquad (7.34)$$

В случае несмещенной оценки параметра, которым мы ограничимся, из (7.32) и (7.33) следует несмещенность $\hat{\eta}(t)$ как оценки $\eta(t)$: $M[e(t)] = 0$. Другими словами, систематической ошибки прогноза нет. Есть только случайная ошибка, дисперсия которой равна

$$\sigma_e^2(t) = M[e^2(t)] = k^2(t)\sigma_{\hat{c}}^2 + \sigma_\xi^2. \qquad (7.35)$$

Если $\xi$ в исходном процессе (7.3) или (7.24) нормальный, то 95%-й интервал для $\eta(t)$ имеет вид $\hat{\eta}(t) \pm 1.96\sigma_e^2$. Эту формулу часто используют как приближенную, даже если закон распределения $\xi$ неизвестен.[17]

Если дисперсия шума и погрешность оценки параметра не велики, то ошибка прогноза (7.35) остается малой, пока мала величина $k(t)$. Последнее имеет место для любого *t*, если $f(t, c)$ не чувствительна к изменениям *c*. В противном случае $k(t)$ и ошибка прогноза могут

---

[17] Вместо дисперсии $\sigma_\xi^2$ можно подставить ее оценку $\hat{\sigma}_\xi^2$. Дисперсии оценок параметров $\sigma_{\hat{c}}^2$ обычно пропорциональны дисперсии шума $\sigma_\xi^2$ и обратно пропорциональны дине ряда *N* или более высоким степеням *N* (см. пп. 7.1.2.1, 7.1.2.2). Приведем формулы для общего случая – вектора параметров **c**. Ковариационную матрицу оценок параметров можно оценить как $Cov(\hat{\mathbf{c}}) = \hat{\sigma}_\xi^2 \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}$, где $A_{j,k} = \sum_{i=1}^{N} \phi_j(t_i)\phi_k(t_i)$ [60]. Диагональные элементы $Cov(\hat{\mathbf{c}})$ – дисперсии оценок параметров: $\hat{\sigma}_{\hat{c}_i}^2 = [Cov(\hat{\mathbf{c}})]_{ii}$. Для нелинейной по параметрам модели оценку ковариационной матрицы получают, обращая матрицу Гессе функции правдоподобия: $Cov(\hat{\mathbf{c}}) = \mathbf{H}^{-1}(\hat{\mathbf{c}})$, где $H_{ij}(\hat{\mathbf{c}}) = \partial^2 \ln L(\hat{\mathbf{c}}) / \partial c_i \partial c_j$.

нарастать со временем. Например, для $f(t,c)=\sin(ct)$ имеем $k(t)=t\cos(c_0 t)$. Это неограниченная функция, поэтому изначально очень малая погрешность прогноза при больших $t$ нарастает до масштаба колебаний наблюдаемой.[18]

Заметим, что величину ошибку прогноза удалось оценить здесь за счет информации о свойствах шума $\xi$. В общем случае «происхождение» ошибок модели неизвестно и оценить их величину, располагая только самой моделью, дающей точечный прогноз, нельзя [251], см. также п. 10.3.

**7.4.1.3. Интерполяция и экстраполяция.** Уже в предыдущем примере видна общая закономерность: гораздо меньшую точность имеет прогноз в удаленные моменты времени, находящиеся за пределами интервала наблюдения (экстраполяция). Это имеет место даже в случае известной структуры модели. Если же структура неизвестна, то для успеха экстраполяции нужно не только точно знать значения параметров, но и знать, что выбранный вид функции $f$ адекватен для описания процесса на широком интервале значений $t$. Последнее чаще не имеет места. Так, алгебраический многочлен высокого порядка, как правило, *очень плохо* экстраполирует.

Тем не менее, в ряде практических случаев прогноз-экстраполяция используется. Например, для прогноза медленных процессов в эконометрике используются модели с $f$ в виде алгебраических многочленов первого или второго порядка. В этих задачах наблюдения (за прибылью предприятий и т.д.) могут проводиться раз в год в течение нескольких лет, а прогноз важен на один год (один шаг) вперед. При этом в силу медленности изменения процесса, более или менее правдоподобным выглядит предположение о том, что линейное или квадратичное изменение сохранится в течение еще одного года. Другими словами, экстраполяция «не далеко вперед» с помощью простых моделей – задача относительно разумная.

---

[18] Именно поэтому дисперсия оценки параметра для процесса (7.19) уменьшается быстрее (как $1/N^3$) с ростом длины ряда, см. п. 7.1.2.2. Поскольку сама функция $f$ чувствительна к изменениям параметра, то и целевая функция $S$ в (7.12) чувствительна к ним. В данном случае квадрат величины $\partial f(t,c)/\partial c\big|_{c=c_0}$ – это информационное количество Фишера, определяющее вклад в функцию правдоподобия каждого нового наблюдения.

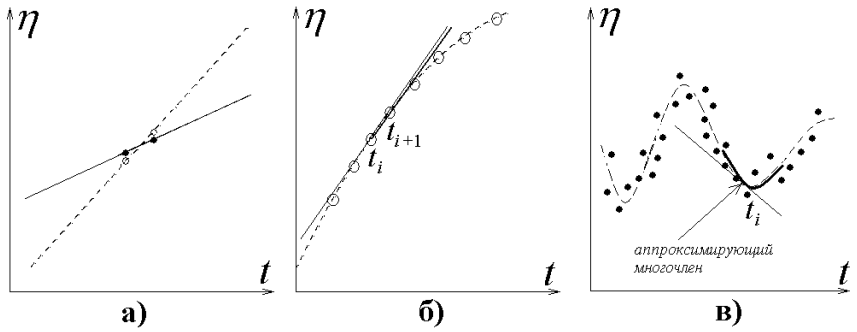



Рис.7.7. а) Прогноз линейной зависимости от времени по двум точкам. Между ними прогноз достаточно точен (интерполяция). В удаленные моменты времени ошибка прогноза велика (экстраполяция). б) Иллюстрация схемы (7.36). Пунктир – график $F(t)$, тонкая линия – истинная касательная, жирная линия – секущая (7.36). в) Дифференцирование с помощью аппроксимирующего многочлена (жирная линия). Тонкая линия – касательная к модельному многочлену, близкая к истинной

Что касается прогноза в промежуточные моменты времени (интерполяции), то он может быть, как правило, осуществлен с достаточной надежностью (рис.7.7,а). Важно лишь предположение о том, что аппроксимируемая зависимость не имеет сильных выбросов между моментами наблюдений. Оно часто вполне разумно. Об интерполяции существует обширная литература [85, 153, 289], в частности, известны следующие методы.

1) Интерполяционные алгебраические многочлены (график проходит через все экспериментальные точки, т.е. используется высокий порядок многочлена) годятся для плавных зависимостей и не очень больших длин ряда (для грубой ориентировки, не более примерно 30 точек).

2) Кусочно-линейная интерполяция (соседние во времени точки попарно соединяются прямыми линиями, получается ломаная) хороша для не слишком гладких зависимостей и высокой частоты выборки.

3) Кубические сплайны (проведение через каждые две соседние точки кубической параболы так, чтобы обеспечить непрерывность первой и второй производных модельной функции) являются весьма эффективным и достаточно универсальным средством приближения гладких зависимостей при высокой частоте выборки [85, 153, 289].

## 7.4.2. Численное дифференцирование

Как уже отмечалось в начале данной главы, интерес могут представлять сами значения параметров модели. Пример – распространенная задача *численного дифференцирования*. Пусть наблюдаемый процесс имеет вид $\eta(t_i) = x(t_i) + \xi(t_i)$, где $t_i = t_0 + i\Delta t$, $x(t_i)$ – гладкая детерминированная функция. Требуется по $\eta(t_i)$ оценить значения производной $dx(t_i)/dt$.

Наиболее распространенный подход – аппроксимация функции $x(t)$

алгебраическим многочленом невысокого порядка в окрестности каждого интересующего нас момента времени $t_i$ и оценка параметров этого многочлена с помощью МНК.[19] Производная модельного многочлена принимается в качестве оценки величины $dx(t_i)/dt$. Если частота выборки достаточно высока и функция $x(t)$ достаточно гладкая, то аппроксимация алгебраическим многочленом обоснована теоремой Тейлора о разложении гладкой функции по степеням отклонения от $t_i$ в окрестности $t_i$.

**7.4.2.1. Дифференцирование при отсутствии шума.** Если $\xi = 0$, обычно используют интерполяционные многочлены. Точность дифференцирования определяется тем, насколько хорошо может быть приближена зависимость $x(t)$ многочленом в выбранной окрестности $t_i$. Наиболее простая схема – провести прямую через точки $t_i$ и $t_{i+1}$ (многочлен 1-го порядка, рис.7.7,б). Оценка производной тогда имеет вид

$$\frac{d\hat{x}(t_i)}{dt} = \frac{\eta_{i+1} - \eta_i}{t_{i+1} - t_i} = \frac{\eta_{i+1} - \eta_i}{\Delta t}. \tag{7.36}$$

Погрешность можно оценить, используя формулу Тейлора $\eta_{i+1} \approx \eta_i + \left(dx(t_i)/dt\right)\Delta t + \left(d^2x(t_i)/dt^2\right)\Delta t^2/2$. Получим ошибку, равную $\left(d^2x(t_i)/dt^2\right)\Delta t/2$, т.е. пропорциональную $\Delta t$. В этом случае говорят, что метод имеет первый порядок точности. Уменьшение $\Delta t$ приводит к более точным оценкам производной, но только при отсутствии шума и погрешностей вычислений. Устремлять $\Delta t$ к нулю нельзя, т.к. при этом проявится некорректность задачи численного дифференцирования (см. ниже).

Можно повысить порядок точности метода, используя многочлены более высокого порядка. Например, строя по точкам $t_{i-1}$, $t_i$ и $t_{i+1}$ многочлен второго порядка, получим формулу

$$d\hat{x}(t_i)/dt = \frac{\eta_{i+1} - \eta_{i-1}}{2\Delta t}. \tag{7.37}$$

Для равномерной выборки она свелась к проведению прямой через точки $t_{i-1}$ и $t_{i+1}$. Ее порядок точности равен 2, т.к. погрешность равна $\left(d^3x(t_i)/dt^3\right)\Delta t^2/6$. Имеются и формулы более высоких порядков [85, 153].

**7.4.2.2. Некорректность задачи.** Покажем, почему нельзя брать очень малую величину $\Delta t$, например, в схеме (7.36). Пусть значения $\eta_i, \eta_{i+1}$ известны с малыми погрешностями $\xi_i, \xi_{i+1}$ порядка $\delta$. Тогда погрешность в

[19] Этот подход имеет много названий, например, фильтр Савицки – Голэя или цифровой сглаживающий многочлен [289].

значении производной составит $(\xi_{i+1} - \xi_i)/\Delta t \sim \delta/\Delta t$. При сколь угодно малой погрешности $\delta$ получим сколь угодно большую погрешность оценки для достаточно малых $\Delta t$. Это означает некорректность постановки задачи – неустойчивость решения по входным данным (см. п. 5.3.2).

Некорректность задачи снимается, если при численном дифференцировании величина порядка $\delta^{1/2}$ используется как минимально возможное значение интервала $\Delta t$. При этом погрешность оценки производной ограничена и стремится к нулю при уменьшении $\delta$. Это – пример *регуляризации* при решении некорректной задачи.

**7.4.2.3. Дифференцирование при наличии шума.** В этом случае непосредственное использование формул типа (7.36) или (7.37) ведет к «усилению шума» и большим ошибкам оценок производных (рис.7.8,а,б). Чтобы снизить влияние шума, модельный алгебраический многочлен строят в некотором достаточно широком окне $[t_{i-k_1\Delta t},\ t_{i+k_2\Delta t}]$ с помощью МНК (рис.7.7,в). Эффективность подхода проиллюстрирована на рис.7.8,в.

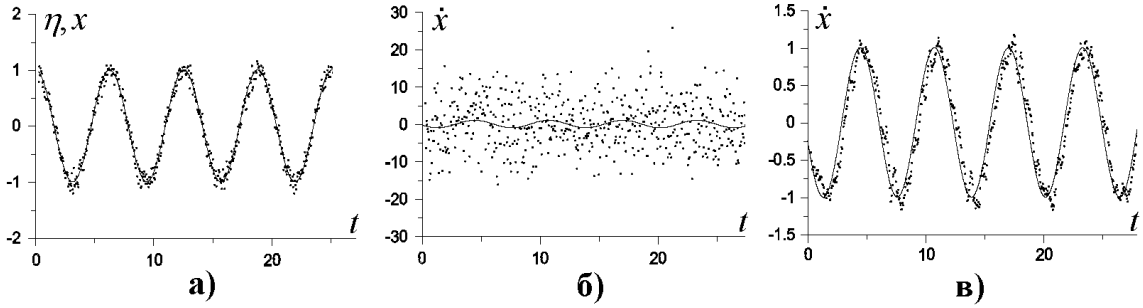



Рис.7.8. Численное дифференцирование $\eta(t) = \cos t + \xi(t)$ при нормальном шуме с дисперсией 0.01 и $\Delta t = 0.01$. а) Исходный ряд, сплошная линия – не зашумленные значения $x(t)$, б) дифференцирование по формуле (7.37), сплошная линия – истинные значения производной, в) дифференцирование с помощью сглаживающего многочлена порядка 2, построенного по 51 точке, сплошная линия – истинные значения производной. Погрешность многократно снижена

Поскольку производная равна одному из коэффициентов многочлена, то величину ее среднеквадратичной погрешности можно получить, рассчитав погрешность оценки этого коэффициента (п. 7.4.1).

Число коэффициентов многочлена должно быть не велико по сравнению с числом точек окна. Чем больше число точек в окне, тем меньше дисперсия оценки (выше точность дифференцирования). Но интервал $\max\{k_1, k_2\}\Delta t$ не должен быть слишком широк, чтобы аппроксимация многочленом низкого порядка оставалась удовлетворительной. Так что при фиксированном порядке многочлена есть некоторая оптимальная ширина окна, когда погрешность из-за шумов и погрешность из-за несовершенства аппроксимации примерно одинаковы. Увеличение порядка многочлена может повысить точность, но только если

при этом удастся использовать значительно более широкое окно. Это зависит от характера функции $x(t)$ и величины $\Delta t$. При фиксированном $\Delta t$ существует некоторое оптимальное значение порядка многочлена для наиболее точного расчета производной. На практике оно составляет, как правило, величину от 1 до 3.

Можно рассчитывать и производные более высоких порядков, используя в качестве оценок соответствующие производные модельного многочлена. Но для оценки производной высокого порядка требуется строить и многочлен высокого порядка, что часто ведет к большим погрешностям. Погрешности дифференцирования быстро растут с ростом порядка производной. В типичных практических постановках удается численно рассчитывать производные не выше третьего порядка (в «хорошем» случае).