

УДК 57.087+519.254+519.213.2

© А. Л. Буляница

МЕТОДЫ ОЦЕНИВАНИЯ ПАРАМЕТРОВ КРИВОЙ ЛОГИСТИЧЕСКОГО РОСТА.

Ч. 1. ОПТИМИЗАЦИЯ УСЛОВИЙ ОЦЕНИВАНИЯ ПРИ НАЛИЧИИ АДДИТИВНОЙ СЛУЧАЙНОЙ ПОМЕХИ

Логистический рост является одной из традиционных математических моделей, применяемых для описания многих популяционных процессов. В частности, динамика количества ДНК при проведении полимеразной цепной реакции в реальном времени (ПЦР-РВ) адекватно описывается подобной моделью. Параметры кривой логистического роста служат основой для выявления необходимой аналитической информации об исходном количестве генетического материала. Объектом исследования является последовательность измерений, каждое из которых содержит в общем случае помимо полезной информативной составляющей случайную центрированную аддитивную помеху и систематическую составляющую в форме линейного тренда первого порядка с априорно неизвестными параметрами. Задачами данного раздела исследования является разработка алгоритмов эффективного оценивания параметров кривой логистического роста при наличии аддитивной центрированной ограниченной случайной помехи.

Кл. сл.: полимеразная цепная реакция (ПЦР), логистический рост, оценивание параметров, случайная помеха, стохастическая аппроксимация

ВВЕДЕНИЕ

Основам осуществления полимеразной цепной реакции (ПЦР) посвящено множество работ, в частности [1–6]. Ранее было доказано, что информативный сигнал в приборах, реализующих полимеразную цепную реакцию в реальном масштабе времени, в первом приближении описывается кривой логистического роста. Даже предложенные альтернативные модели трехстадийной динамики (практическое отсутствие роста количества ДНК, экспоненциальный (геометрический) рост и выход на стадию насыщения) [7, 8] не противоречат на качественном уровне гипотезе о логистическом росте.

Одной из принятых форм этой кривой является

$$X_n = \frac{X^*}{1 + (B - 1)a_0^{-n}}. \quad (1)$$

Здесь X_n — величина сигнала на n -м шаге измерения ($n = 0, 1, \dots, N$, где N обычно 30–40).

Подобная форма зависимости получается с помощью квантования по времени (дискретизации) решения задачи Коши

$$dx/dt = \lambda x(1 - x/x^*); \quad x(0) = x_0. \quad (2)$$

Соответствующее дифференциальное уравнение первого порядка есть уравнение Бернулли (см., например, [9]) Оно решается в квадратурах отно-

сительно $y = x^{-1}(t)$ при любых зависимостях от времени параметров $\lambda = \lambda(t)$; $x^* = x^*(t)$. В частности, в работе [10] рассмотрены формы решений уравнения (2) при линейных зависимостях параметров уравнения от времени.

Определяющими параметрами являются следующие.

1) Коэффициент a_0 , обычно находящийся в пределах 1.78–1.98 (при теоретическом пределе, равном 2), характеризующий эффективность проведения реакции. Этот коэффициент определяет средний прирост количества ДНК за 1 цикл.

2) Величина X^* характеризует максимальный информативный сигнал, т. е. либо величину сигнала при максимальном количестве ДНК (при достижении режима насыщения), либо максимально допустимую величину сигнала, определяемую чувствительностью измерения, разрядностью устройств (в частности, АЦП) и другими факторами, связанными не с ПЦР, а с выбранным режимом измерения.

3) Параметр B численно равен отношению X^* и X_0 — сигнала, соответствующего начальному (достаточно малому) количеству ДНК. Соответствующее отношение B обычно не менее 10^3 , но может достигать 10^6 – 10^7 .

Основные особенности рассматриваемых информативных сигналов.

1) Наличие аддитивной помехи, которую будем полагать ограниченной в диапазоне $\pm\delta$ (возможно, от единиц до 10–20 отсчетов).

2) Максимальный сигнал X^* полагаем превосходящим 1000 (возможно, до 3000–4000 отсчетов). Эта величина, в частности, определяется разрядностью используемого АЦП.

3) Сигнал, соответствующий начальному количеству ДНК X_0 , как правило, не превосходит 1 отсчета (возможно, 10^{-3} и менее).

Главные методические проблемы анализа информативных сигналов состоят в следующем:

а) значимые параметры X^* и a_0 априорно неизвестны, а величина X_0 по существу и является искомым аналитической информацией;

б) при определенных условиях анализа кривая может не достигать стадии насыщения;

в) калибровка прибора обычно осуществляется на основе наиболее устойчивой комплексной характеристики сигнала — так называемого порогового цикла, т. е. стадии анализа, когда информативный сигнал обладает наибольшей линейностью. Аналитически положение порогового цикла может быть вычислено из соотношения $n^* = \ln(B-1) / \ln(a_0)$. При этом следует учесть, что n^* необходим для определения B . Например, если при $a_0 = 2$ пороговый цикл равен 20.00, то этому значению B при $a_0 = 1.8$ будет соответствовать сильное смещение положения порогового цикла до 23.58. Таким образом, при осуществлении калибровки по положению порогового цикла n^* необходимо учитывать и эффективность ПЦР (т. е. величину a_0).

Дальнейшие рассуждения касаются реализации алгоритмов последовательного оценивания базовых параметров информативного сигнала (a_0 , X^* , а также X_0 или/и n^*).

НАЧАЛЬНЫЕ УСТАНОВКИ АЛГОРИТМА. ОСНОВНЫЕ ПРИНЦИПЫ ОЦЕНИВАНИЯ ИНФОРМАТИВНЫХ ПАРАМЕТРОВ — ПОСЛЕДОВАТЕЛЬНОСТЬ, РЕАЛИЗАЦИЯ И КАЧЕСТВО ОЦЕНИВАНИЯ

Полагаем, что информативный сигнал X_n содержит аддитивную ограниченную помеху ζ_n . (Для конкретизации считаем помеху равномерной и независимой).

Исследуем вторую разность информативного сигнала:

$$2X_n - X_{n-1} - X_{n+1} = 2\zeta_n - \zeta_{n-1} - \zeta_{n+1} + X^*(a_0 - 2 + 1/a_0) \frac{t(1-t)}{(1+t)(1+t/a_0)(1+a_0t)}. \quad (3)$$

Здесь $t = (B-1)/a_0^n$. Исследование экстремумов

второй разности (и связанной с ним кривизны) зависимости (3) дает следующие результаты: $t = 1$ — точка нулевой кривизны (соответствует наиболее линейному участку кривой и положению порогового цикла $n = n^*$); в зависимости от a_0 максимальная кривизна кривой реализуется при t , близких к 4 и $1/4$ (точнее: при $a_0 = 2$ эти значения 4.013 и 0.249, при $a_0 = 1.9$ и менее — 3.87 и 0.26). Практически эти точки отстоят примерно на 2 отсчета от точки перегиба (или порогового цикла).

Ориентировочно оценим максимальную вторую разность в этих и соседних точках. Зададим $a_0 = 2$, т. к. при достаточно близких значениях параметра результаты будут количественно схожи. Следует отметить, что замена t на $1/t$ означает движение по логистической кривой в другую сторону от точки перегиба. Соответствующие величины второй разности в этом случае будут совпадать по модулю и иметь противоположный знак. При $t > 1$ эти отсчеты предшествуют точке перегиба, и вторая разность отрицательна; при $t < 1$ рассматриваются отсчеты после точки перегиба или порогового цикла.

Оценим величину второй разности при отсчетах, соответствующих $t = 8, 4, 2$, т. е. в трех предшествующих пороговому циклу отсчетах (полагая при этом, что n^* — целая величина). При дробных n^* характерные величины второй разности в трех ближайших к пороговому циклу отсчетах будут соответствовать названным величинам t , деленным на a_0 в степени $\{n^*\}$ — дробная часть от n^* — (т. е. на величину от 1 до $\sqrt{2}$).

В отсчетах $t = 8, t = 4, t = 2$ вторая разность — по абсолютной величине $X^*/27.32, X^*/22.5$ и $X^*/30.0$ соответственно. Суммарная аддитивная помеха при формировании второй разности (3) достоверно не превзойдет по величине 4δ . Тем самым гарантировано, что при условии $\delta < X^*/120$ по крайней мере в 3 последовательных отсчетах, предшествующих точке перегиба, будет сохраняться знак второй разности. Далее при переходе через точку перегиба произойдет смена знака второй разности с отрицательного на положительный. В указанной ситуации с точностью примерно 0.5 отсчета может быть установлено положение точки перегиба (по интервалу смены знака второй разности).

Более подробно исследуем аддитивную помеху для второй разности. В первом приближении закон распределения помехи каждого измерения X_n полагаем равномерным. В принципе допустимо рассматривать любую гипотезу о распределении случайной составляющей помех при наложенных ограничениях о симметричности (и как следствие центрированности) и ограниченности. При независимости помех на $(n-1)$, n и $(n+1)$ отсчетах плотность распределения вероятностей (ПРВ) результирующей помехи получается методом сверт-

ки. Удобнее использовать аппарат характеристических функций (см., например, [11]). Переход к характеристическим функциям аналогичен преобразованию Лапласа с заменой аргумента p на мнимый аргумент (ju).

Для равномерной в диапазоне $[-\delta, \delta]$ помехи характеристическая функция имеет вид

$$\Theta^*(ju) = \frac{1}{2\delta} \frac{e^{ju\delta} - e^{-ju\delta}}{ju}. \quad (4)$$

Результующей помехе соответствует произведение характеристических функций вида (4), что приводит к выражению:

$$\Theta(ju) = \frac{1}{16\delta^3} \frac{(e^{ju2\delta} - e^{-ju2\delta})(e^{ju\delta} - e^{-ju\delta})^2}{(ju)^3}. \quad (5)$$

Здесь первый множитель соответствует слагаемому $2\zeta_n$, два других множителя в числителе соответствуют помехам ζ_{n-1} и ζ_{n+1} .

Обращение характеристической функции (5) аналогично переходу к оригиналу для преобразования Лапласа позволяет получить выражения для ПРВ в форме

$$p(x) = \frac{1}{32\delta^3} [(x+4\delta)^2 \eta(x+4\delta) - 2(x+2\delta)^2 \eta(x+2\delta) + 2(x-2\delta)^2 \eta(x-2\delta) + (x-4\delta)^2 \eta(x-4\delta)].$$

Здесь $\eta(x)$ — функция Хевисайда (функция единичного скачка). Обращение характеристической функции можно выполнить либо с использованием таблиц преобразования Лапласа [12], либо на основе теории вычетов. Соответственно для характерных интервалов явное выражение ПРВ примет вид:

$$\begin{aligned} x < -4\delta: & \quad p(x) = 0, \\ x \in [-4\delta, -2\delta]: & \quad p(x) = (x+4\delta)^2 / (32\delta^3), \\ x \in [-2\delta; 0]: & \quad p(x) = (8\delta^2 - x^2) / (32\delta^3); \\ \forall x \Rightarrow & \quad p(-x) = p(x). \end{aligned} \quad (6)$$

График ПРВ представлен на рис. 1, а соответствующая функция распределения иллюстрируется рис. 2.

На основе (6) легко рассчитать вероятности попадания суммарной помехи в заданный диапазон значений: вероятность попадания в интервал $[0, 1\delta]$ равна $23/96$, $[1\delta, 2\delta]$ — $17/96$, $[2\delta, 3\delta]$ — $7/96$ и $[3\delta, 4\delta]$ — $1/96$. Математическое ожидание равно нулю, дисперсия как сумма дисперсий 3 независимых слагаемых будет $(4+1+1)\delta^2/3$. Таким образом видно, что подобная суммарная помеха "фокусируется" в окрестности $x = 0$.

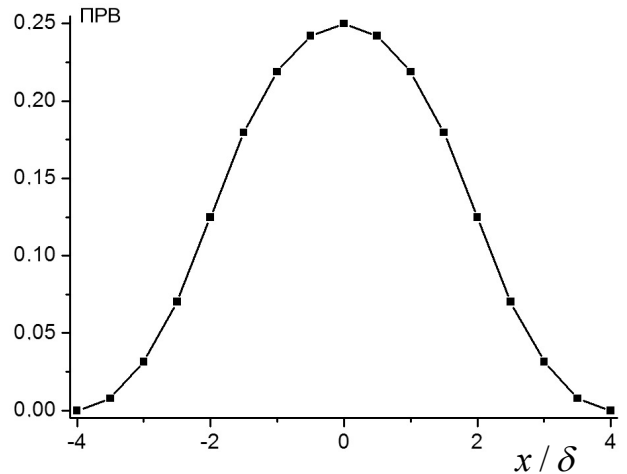


Рис. 1. ПРВ случайной составляющей второго разностного сигнала

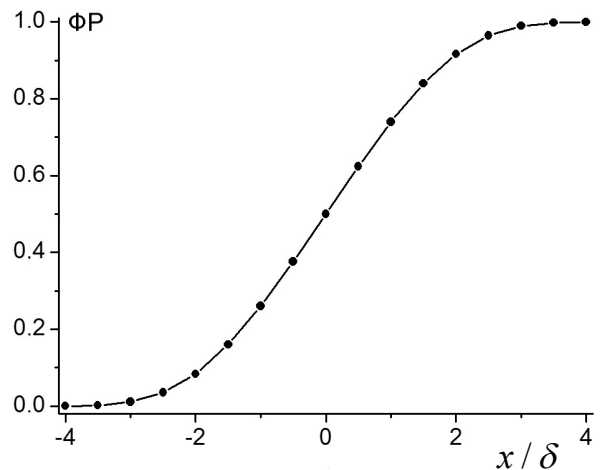


Рис. 2. Функция распределения (ФР) случайной составляющей второго разностного сигнала

Алгоритмы оценивания информативных параметров сигнала a_0, X^* и B базируются на принципе построения постоянного сигнала (тренда нулевого порядка) на основе имеющихся измерений X_n при аддитивной случайной помехе ζ_n .

На первом этапе строится последовательность оценок величины a_0 в форме

$$\hat{a}_0 = \frac{X_{n+2}(X_{n+1} - X_n)}{X_n(X_{n+2} - X_{n+1})}. \quad (7)$$

Указанная схема, учитывающая три последовательных измерения, начиная с отсчета n (далее трехточечная схема), обладает большей алгоритмической простотой, но меньшей точностью. В принципе возможно решение задачи оптимизации выбора базовых отсчетов $n, n + 1$ и $n + 2$ по критерию

рию минимума дисперсии ошибки определения a_0 . Соответствующее положение отсчетов должно удовлетворять условию $t \approx 1.7$. (Это положение точки оптимума достаточно слабо зависит от измеряемой величины a_0). На практике следует учесть, что этот оптимум может реализоваться приближенно, поскольку n должно быть целым.

Более точной схемой для построения оценки величины a_0^k и как следствие a_0 является 4-точечная схема с использованием сигнала в отсчетах $n, n+1, n+k$ и $n+k+1$:

$$\hat{a}_0^k = \frac{X_{n+k} X_{n+k+1} (X_{n+1} - X_n)}{X_n X_{n+1} (X_{n+k+1} - X_{n+k})}. \quad (8)$$

Таким образом, используются разделенные на k отсчетов две пары последовательных измерений.

Дисперсия ошибки оценивания (в линейном приближении) зависит от n и k (задача оптимизации по критерию минимума дисперсии решается поиском нетривиального минимума функции 2 переменных t и k). Положение оптимума также слабо зависит от величины a_0 : при a_0 достаточно больших (1.90 и более) положение оптимума соответствует $t \approx 4.5, k = 4$; при менее эффективной ПЦР ($a_0 \approx 1.80$) оптимальными будут условия $t \approx 5$ и $k = 5$. В принципе незначительное отклонение базовых отсчетов ($n, n+1, n+k$ и $n+k+1$) от оптимальных слабо влияет на величину дисперсии ошибки оценивания a_0 .

Вторым этапом является определение максимального сигнала X^* в соответствии с выражением

$$\hat{X}^* = \frac{\hat{a}_0^k - 1}{\hat{a}_0^k / X_{n+k} - 1 / X_n}. \quad (9)$$

После вычисления X^* имеется возможность оценить $\ln(B-1)$, далее вычислить B или/и положение порогового цикла n^* . Используются уже выбранные отсчеты со значениями величин сигнала X_n :

$$\ln(\hat{B}-1) = \ln(\hat{X}^* / X_n - 1) + n \ln(\hat{a}_0). \quad (10)$$

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ И ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ ПРОЦЕДУРЫ НАХОЖДЕНИЯ ПАРАМЕТРОВ КРИВЫХ ЛОГИСТИЧЕСКОГО РОСТА

В качестве генератора равномерно распределенной случайной величины использована программа RANDOM на языке C++. Генерация соответствующей случайной величины с диапазоном $[-\delta, \delta]$ выполнена как

$$2\delta(\text{RANDOM}(1001)/1000 - 1/2).$$

В круглых скобках показан способ генерации равномерно распределенной центрированной случайной величины в диапазоне $[-0.5, 0.5]$.

Далее представлены результаты имитационного моделирования расчета параметров информативного сигнала. Заданы следующие параметры информативного сигнала и случайной помехи: $a_0 = 1.920$; $X^* = 3500$; $X_0 = 0.02$; $\delta = 10$. Этим параметрам соответствует $n^* = 18.507$; $B = 174999$.

По критерию смены знака второй разности выявлено положение порогового цикла на интервале 18–19 отсчетов. Выбрана 4-точечная схема (8) для оценивания a_0 : $k = 4$ и $n = 13$ –18. Значению $n = 13$ соответствует $t = 41.5$ (исходя из предположения, что $a_0 = 1.9$), т. е. этот отсчет примерно на 3–4 такта предшествует оптимальному режиму оценивания a_0 .

6 оценок a_0 равны: 1.918, 1.896, 1.918, 1.906, 1.917, 1.927. Результирующая оценка 1.917. На основе этих же измерений получены, согласно (9), оценки X^* : 3395, 3385, 3491, 3490, 3504, 3505, и итоговая оценка 3490.2, а также оценки (10) величины $\ln(B-1)$: 12.029, 12.031, 12.042, 12.045, 12.038, 12.051. На базе результирующей оценки $\ln(B-1)$ получено $B = 169388$ и как следствие — $X_0 = 0.0206$; $n^* = 18.497$. Все оценки значения достаточно близки к заданным значениям.

Оценки параметров вычислялись на основе различных базовых измерений вблизи области оптимальных значений (n, k). В качестве результирующей оценки была использована медиана выборки 6 измерений, т. е. полусумма 2 центральных порядковых статистик. Как известно, она обладает робастностью и одновременно с этим учитывает и различие оценок, полученных на основе разных отсчетов. Далее будет рассмотрена альтернативная процедура получения результирующей оценки.

В целом наблюдается хорошая сходимость оценок параметров к истинным значениям. Условие приближенного "отслеживания" точки перегиба в форме $\delta < X^*/120$ было выполнено во всех случаях более чем с двукратным запасом.

АНАЛИЗ ПОГРЕШНОСТЕЙ ОЦЕНИВАНИЯ ПАРАМЕТРОВ КРИВОЙ ЛОГИСТИЧЕСКОГО РОСТА

Общий вид оценки — $\hat{a}_0 = a_0 + \Delta a = a_0(1 + \varepsilon_a)$ при Δa и ε_a — абсолютной и относительной погрешностях. В первом приближении (приближении дифференциала первого порядка) для 3-точечной схемы (7) абсолютная погрешность имеет вид:

$$\Delta a \approx \frac{a_0}{X^*} (K_n \zeta_n + K_{n+1} \zeta_{n+1} + K_{n+2} \zeta_{n+2}),$$

где

$$\begin{aligned} K_n &= -\frac{a_0}{a_0 - 1}(t + 2 + 1/t), \\ K_{n+1} &= \frac{a_0 + 1}{a_0 - 1}(t/a_0 + 2 + a_0/t), \\ K_{n+2} &= -\frac{1}{a_0 - 1}(t/a_0^2 + 2 + a_0^2/t). \end{aligned} \quad (11)$$

Учитывая характер аддитивных случайных погрешностей, в частности их центрированность и независимость, можно сделать выводы:

- а) оценка — несмещенная;
- б) ее дисперсия равна

$$\sigma_a^2 = \frac{a_0^2}{(X^*)^2} \delta^2 \sigma_0^2 (K_n^2 + K_{n+1}^2 + K_{n+2}^2). \quad (12)$$

Здесь σ_0^2 — дисперсия исходной помехи, приведенная к диапазону $[-1, 1]$. В случае равномерной помехи $\sigma_0^2 = 1/3$. Выражение в круглых скобках является функцией t , и возможны постановка и решение задачи поиска минимума соответствующей функции. Как говорилось ранее, имеется нетривиальное решение $t \approx 1.7$, слабо зависящее от a_0 ($t \approx 1.73$ при $a_0 = 2$, $t \approx 1.68$ при $a_0 = 1.9$, $t \approx 1.61$ при $a_0 = 1.8$). В частности, при $a_0 = 2$ минимальная дисперсия ошибки оценивания

$$\sigma_a^2 \approx \frac{4 \cdot 242.268}{3} (\delta / X^*)^2 \approx 323 (\delta / X^*)^2.$$

В общем случае при произвольных истинных значениях параметров кривой оптимуму может соответствовать нецелое значение n , и потребуются выбрать ближайший к точке оптимума отсчет n .

Схожий вид (и аналогичные свойства) имеет дисперсия ошибки определения a_0 при 4-точечной схеме (8):

$$\sigma_a^2 = \frac{a_0^2}{(X^*)^2} \delta^2 \sigma_0^2 \frac{1}{k^2} (K_n^2 + K_{n+1}^2 + K_{n+k}^2 + K_{n+k+1}^2), \quad (13)$$

где

$$\begin{aligned} K_n &= -\frac{a_0}{a_0 - 1}(t + 2 + 1/t), \\ K_{n+1} &= \frac{1}{a_0 - 1}(t/a_0 + 2 + a_0/t), \\ K_{n+k} &= \frac{a_0}{a_0 - 1}(t/a_0^k + 2 + a_0^k/t), \\ K_{n+k+1} &= -\frac{1}{a_0 - 1}(t/a_0^{k+1} + 2 + a_0^{k+1}/t). \end{aligned}$$

Правые множители формируют функцию 2 пе-

ременных (t, k), и возможен поиск минимума указанной функции. Точка минимума при заданном a_0 искалась приближенным методом спирального (координатного) спуска. Как говорилось ранее, координаты этой точки слабо зависят от a_0 . При эффективно реализованной ПЦР (т. е. при $a_0 \approx 2$) точка оптимума лежит вблизи $t = 4.7, k = 4$. В этом случае минимальная дисперсия ошибки оценивания определяется как $\sigma_a^2 \approx \frac{4 \cdot 26.500}{3} (\delta / X^*)^2 \approx 35.3 (\delta / X^*)^2$. Тем самым стандартное отклонение оценки (13) в случае 4-точечной оптимальной схемы будет приблизительно в 3 раза меньшим по сравнению с (12), т. е. аналогичной 3-точечной схемой.

Проведенное имитационное моделирование, включающее генерацию идеальной кривой логистического роста и аддитивной помехи с заданным законом распределения (100–250 различных реализаций), подтверждает указанные выше закономерности оценки a_0 . В частности, близкая к оптимальной 4-точечная схема при заданных условиях ($a_0 = 2.000; X^* = 3000; X_0 = 10; \delta = 5$) дает следующие характеристики оценки a_0 : математическое ожидание 1.99939, среднее квадратичное отклонение 0.01023 (в соответствии с расчетной формулой (13), среднее квадратичное отклонение в указанных условиях должно составить 0.009907). Некоторое расхождение может быть объяснено неидеальностью характеристик датчика равномерной случайной величины. В частности, не в полной мере обеспечивается несмещенность случайной величины, имеются незначительные асимметрия и превышение коэффициента эксцесса.

При анализе ошибки оценивания X^* , связанной как с наличием аддитивной погрешности в базовых измерениях X_n и X_{n+k} , так и с погрешностью оценивания a_0 , следует учитывать, что погрешности могут не быть независимыми, если оценивание a_0 и X^* базируется на одних и тех же измерениях (на отсчетах n и $n+k$). Представляет интерес решение задачи о выборе точек для оценивания X^* : что является более выгодным по точности — использование общих точек или различных? Решение этой задачи будет приведено далее.

В первом приближении приращение функции нескольких переменных можно вычислить на основе аналога дифференциала первого порядка:

$$\Delta X^* = \frac{\partial X^*}{\partial a_0} \Delta a_0 + \frac{\partial X^*}{\partial X_n} \Delta X_n + \frac{\partial X^*}{\partial X_{n+m}} \Delta X_{n+m} \quad \text{при ис-}$$

пользовании оценки $\hat{X}^* = \frac{\hat{a}_0^m - 1}{\hat{a}_0^m / X_{n+m} - 1 / X_n}$, ана-

логичной (9). Здесь приращения ΔX_n и ΔX_{n+m} равны соответственно ζ_n и ζ_{n+m} . (Номера n и $n+m$ могут не совпадать с аналогичными номерами отсчетов,

по которым осуществлялось оценивание a_0).

$$\text{Задавая } t^* = (B-1)/a_0^n, \text{ получим, что } \frac{\partial X^*}{\partial a_0} = -\frac{m}{a_0^m - 1} \frac{X^* t^*}{a_0}, \frac{\partial X^*}{\partial X_n} = -\frac{1}{a_0^m - 1} (1+t^*) \text{ и } \frac{\partial X^*}{\partial X_{n+m}} = \frac{a_0^m}{a_0^m - 1} (1+t^*/a_0^m).$$

В случае несовпадения отсчетов, по которым осуществлялось оценивание a_0 и X^* , все три слагаемых в ΔX^* будут независимы, и соответствующие дисперсии оценок сложатся:

$$\sigma_{X^*}^2 = \frac{1}{(a_0^m - 1)^2} \left[\frac{m^2}{a_0^2} (X^*)^2 (t^*)^2 \sigma_{a_0}^2 + a_0^{2m} (1+t^*/a_0^m)^2 \delta^2 \sigma_0^2 + (1+t^*)^2 \delta^2 \sigma_0^2 \right].$$

У этой функции имеется тривиальный асимптотический минимум $t^* \rightarrow 0$. Практически это означает, что наиболее точные измерения X^* выполняются при максимальных номерах отсчетов n и $n+m$. Таким образом, для нескольких последних измерений требуется вычислить t^* ; в случае его малости дисперсия ошибки оценивания X^* будет примерно равна $\sigma_{X^*}^2 \approx \frac{a_0^{2m} + 1}{(a_0^m - 1)^2} \delta^2 \sigma_0^2$, т. е. лишь незначительно превзойдет дисперсию аддитивной помехи, равную $\delta^2 \sigma_0^2$.

Другим подходом является оценивание X^* на основе тех же измерений, что и при вычислении оценки a_0 , т. е. $t^* = t$, $m = k$ для 4-точечной схемы (8). При этом для оценивания a_0 и X^* (по (9)) используются одни и те же отсчеты n , $n+1$, $n+k$ и $n+k+1$ и

$$\Delta X^* \approx \frac{1}{a_0^k - 1} (K_n^* \zeta_n + K_{n+1}^* \zeta_{n+1} + K_{n+k}^* \zeta_{n+k} + K_{n+k+1}^* \zeta_{n+k+1}),$$

где

$$K_n^* = \frac{a_0 t}{a_0 - 1} (t + 2 + 1/t) - (1+t),$$

$$K_{n+1}^* = -\frac{t}{a_0 - 1} (t/a_0 + 2 + a_0/t),$$

$$K_{n+k}^* = -\frac{a_0 t}{a_0 - 1} (t/a_0^k + 2 + a_0^k/t) + a_0^k + t,$$

$$K_{n+k+1}^* = \frac{a_0 t}{a_0 - 1} (t/a_0^{k+1} + 2 + a_0^{k+1}/t).$$

Поскольку все четыре слагаемых являются независимыми, то их дисперсии складываются при вычислении дисперсии оценивания X^* . Наиболь-

шая по абсолютной величине ошибка оценивания X^* определяется полуразмахом помехи δ , а также абсолютными значениями коэффициентов K_i^* . Если значение t находится достаточно близко к 4 (т. е. вблизи оптимума для вычисления a_0), то максимальная погрешность оценивания X^* теоретически может составить около 12δ , т. е. огромную величину (хотя и вероятность этого ничтожно мала).

Вывод. Многократно более эффективной является следующая реализация алгоритма:

– на первом этапе при режимах, близких к оптимальным, производится оценивание параметра a_0 ;

– на втором этапе по набору измерений, достаточно близких к участку насыщения (т. е. по заключительным имеющимся отсчетам кривой логистического роста, соответствующим достаточно малым t^*), производится, согласно (9), оценивание другого параметра X^* . Эффективность подобного подхода при оценивании X^* иллюстрируется далее.

Пусть, для примера, сгенерирована последовательность 40 отсчетов с параметрами $a_0 = 1.910$; $X^* = 3000$; $X_0 = 0.50$ и добавлена равномерная центрированная помеха с $\delta = 10$. Далее по (8) с $k = 4$ и $n = 10-15$ проведено оценивание a_0 . После этого выполнен расчет X^* по тем же отсчетам. Получены следующие оценки: 3063, 2997, 2982, 3002, 3001, 3009 (стандартное отклонение $\sigma \approx 29.6$). Для сравнения по измерениям $n = 29-34$ последовательность таких же оценок (9) для X^* имеет вид: 2990, 3010, 2994, 3006, 3004, 3004 ($\sigma \approx 7.8$). Таким образом, оценивание X^* по более поздним отсчетам является существенно более эффективным, исходя из величины среднеквадратичного отклонения ошибки оценивания. Аналогичная тенденция наблюдается и при иных параметрах a_0 , X^* , X_0 и диапазоне помехи.

Заключительным этапом является оценивание параметра B и на его основе величины порогового цикла n как точки перегиба кривой логистического роста. Используется формула (10) и строятся выборки значений $F = \ln(\hat{B} - 1) = \ln(\hat{X}^* / X_n - 1) + n \ln(\hat{a}_0)$. Тогда погрешность оценивания величины F в приближении первого дифференциала будет $\Delta F = \frac{\partial F}{\partial a_0} \Delta a_0 + \frac{\partial F}{\partial X^*} \Delta X^* + \frac{\partial F}{\partial X_n} \Delta X_n$. Полагаем, что предшествующие параметры были оценены ранее в оптимальных условиях и, следовательно, по различным отсчетам. Тем самым соответствующие погрешности оценивания были независимы, и дисперсии могут складываться.

Частные производные равны соответственно

$$\frac{\partial F}{\partial a_0} = n/a_0, \quad \frac{\partial F}{\partial X^*} = \frac{1}{X^* - X_n} \quad \text{и} \quad \frac{\partial F}{\partial X_n} =$$

$= \frac{1}{X^*/X_n - 1} \frac{(-X^*)}{X_n^2} = -\frac{X^*}{X_n(X^* - X_n)}$. Полагая, что измерение $X_n = X^*/(1+t')$, где t' не совпадает ни со значениями t (при условиях оптимального вычисления a_0), ни со значениями t^* (при оптимальном вычислении X^*) получим следующее.

а) Дисперсии a_0 и X^* при оптимальных режимах оценивания равны соответственно $106\delta^2\sigma_0^2/(X^*)^2$ и $\delta^2\sigma_0^2$ (т. к. дисперсия ошибки оценивания X^* в оптимальных условиях оценивания лишь незначительно превосходит дисперсию аддитивной помехи любого из измерений).

б) В приведенных выше обозначениях $\frac{\partial F}{\partial X^*} = \frac{1}{X^*} \frac{1+t'}{t'}$, $\frac{\partial F}{\partial X_n} = -\frac{(1+t')^2}{X^*t'}$ и $n = (\ln(B-1) - \ln(t'))/\ln(a_0)$.

в) Тогда дисперсия ошибки оценивания F примерно равна $\frac{\delta^2\sigma_0^2}{(X^*)^2} \left[\frac{(1+t')^4}{t'^2} + \frac{(1+t')^2}{t'^2} + 106 \frac{n^2}{a_0^2} \right]$.

В принципе возможен поиск минимума дисперсии по параметру t' , который, безусловно, зависит от истинного значения B . Необходимое условие экстремума при этих данных сведется к поиску положительного решения трансцендентного уравнения

$$\begin{aligned} & \frac{212}{a_0^2 \ln^2(a_0)} \ln(t') - 4/t'^2 - 6/t' + 4t' + 2t'^2 = \\ & = \frac{212 \ln(B-1)}{a_0^2 \ln^2(a_0)}. \end{aligned} \quad (14)$$

В случае нахождения такого корня $t' > 1$, соответствующая вторая производная, безусловно, положительна, что свидетельствует об экстремуме типа минимума.

В частности, при $B \approx 1000$ и $a_0 = 2$, оптимум (14) находится близко от $t' = 15$; соответствующее стандартное отклонение оценки логарифма (т. е. функции F) примерно равно $20.5\delta/X^*$ (при $\delta = 10$, $X^* = 3000$ отклонение логарифма на соответствующую величину 0.0683 приведет к отклонению оценки B в пределах 936–1071).

АЛГОРИТМ ПОСТРОЕНИЯ РЕЗУЛЬТИРУЮЩЕЙ ОЦЕНКИ ИНФОРМАТИВНЫХ ПАРАМЕТРОВ НА ОСНОВЕ КОРОТКОЙ ВЫБОРКИ

Ранее в качестве такого алгоритма предлагалось использование медианы по четному числу отсчетов. Практически аналогичного результата можно достичь при применении алгоритма стохастической аппроксимации Роббинса—Монро.

Наиболее удобной представляется робастная модификация Фабиана—Цыпкина. Основанием для использования указанного алгоритма служит модель сигнала (выборка оценок, представляющая собой постоянную величину — истинное значение с аддитивной помехой с априорно неизвестным законом распределения).

Общие принципы организации алгоритма оценивания достаточно детально рассмотрены в работах [13, 14]. Выполняется рекурсивное построение оценок по правилу:

$$c_{n+1} = c_n - \beta/n \cdot \psi(c_n - x_{n+1}). \quad (15)$$

$$\text{При этом } \psi(z) = \begin{cases} -1, & \text{для } z \leq -\Delta, \\ 0, & |z| < \Delta, \\ 1, & z \geq \Delta; \end{cases}$$

$x_{n+1} = c^* + \xi_{n+1}$ — измерение; ξ_{n+1} — помеха; c_n, c_{n+1} — оценки величины c^* на n -м и $n+1$ -м шагах оценивания; β — параметр алгоритма; 2Δ — величина зоны нечувствительности. В предельном случае (15) при $\Delta = 0$ получается сигнатурный алгоритм Я.З. Цыпкина, т. к. $\Psi(z) = \text{sign}(z)$. Начальное приближение c_1 оценок a_0, X^* и X_0 может совпадать с первой из полученных оценок. При оценивании a_0 в качестве первого приближения можно выбрать и априорную оценку 1.90, соответствующую средней эффективности ПЦР.

Степень близости оценки к истинному значению можно определить в соответствии с [13, 14] на основе анализа знака "поправок", т. е. при наличии немонотонности последовательности оценок. В противном случае при монотонном увеличении (уменьшении) оценок достаточно вероятно расхождение оценки и истинного значения параметра. При оценивании X^* и X_0 априорная информация о возможном значении этих параметров отсутствует, и в качестве первого приближения следует брать первую оценку величины.

Параметр β в (15) соизмерим с дисперсией соответствующей оценки: при оценивании X^* сопоставим с величиной δ (т. е. 10–15 квантов), при оценивании a_0 дисперсия оценки близка к $6\delta/X^*$ (т. е. около 0.03). Соответствующий выбор β требует использования значений 0.05–0.10. Наконец, при оценивании величины $F = \ln(B-1)$ стандартное отклонение примерно равно $20\delta/X^*$, т. е. около 0.1. Тем самым следует выбрать β порядка 0.2. Выбранная величина зоны нечувствительности 2Δ должна быть сопоставима с требуемой точностью определения параметра. С практической точки зрения достаточным является оценивание a_0 с точностью до 0.01, X^* с точностью до 1 или даже нескольких квантов, точность при оценивании F определяет с масштабным множителем $1/\ln(a_0) \approx 1.4\div 1.7$ погрешность нахождения положения порогового цикла n^* .

Выборки измерений параметров и оценок, полученных по алгоритму (15)

β	Δ	Параметр	Измерения / оценки *						Итоговая оценка
0.03	0.001	a_0	1.918	1.896	1.918	1.906	1.917	1.927	1.917
			1.918	1.888	1.903	1.903	1.911	1.917	
10	0.5	X^*	3492	3511	3494	3506	3504	3504	3504.8
			3492	3502	3497	3500.3	3502.8	3504.8	
0.5	0.01	F	12.029	12.031	12.042	12.045	12.038	12.051	12.05
			12.03	12.03	12.08	12.05	12.03	12.05	

* Измерения в верхней строке, оценки — в нижней.

Представляется достаточным нахождение n^* с точностью до 0.2–0.3 цикла (отсчета). Тогда получаем результирующие оценки параметров на базе процедуры стохастической аппроксимации (в таблице). По правилу знаков "поправок", эти оценки близки к истинному значению. Исходные оценки, приведенные в таблице, получены при условиях, близких к оптимальным. Результирующие оценки: $a_0 = 1.917$, $X^* = 3504.8$, $F = 12.05$, из чего следует, что $B \approx 171100$, $X_0 \approx 0.0205$ и $n^* \approx 18.52$. Они весьма близки к исходным данным, приведенным в разделе "Имитационное моделирование...".

ЗАКЛЮЧЕНИЕ

В работе исследованы различные алгоритмы оценивания параметров кривой логистического роста. Объединяющим принципом является переход к построению оценок в форме постоянного сигнала (тренда нулевого порядка) с аддитивной случайной помехой. Были проанализированы возможные алгоритмы построения подобных оценок, описаны законы формирования случайной погрешности оценки с доказательством центрированности соответствующих плотностей распределения вероятностей, а также решены задачи оптимизации выбора базовых отсчетов по критерию минимизации дисперсии ошибок оценивания. Учет влияния систематической составляющей сигнала планируется осуществить в ч. 2 статьи.

Данная работа выполнялась при финансовой поддержке в рамках программ и проектов: Аналитической ведомственной целевой программы "Развитие научного потенциала высшей школы" — проект "Исследования и диагностика клеточных структур: новые методические подходы и инстру-

ментальные решения на основе сканирующей зондовой микроскопии и микрочиповых технологий" (№ 4247); Программы фундаментальных исследований Президиума РАН № 20 "Создание и совершенствование методов химического анализа и исследования структуры веществ и материалов" — проект "Микрочиповые аналитические системы для метода молекулярных колоний"; Программы фундаментальных исследований Президиума РАН на 2009 г. № 27 "Основы фундаментальных исследований нанотехнологий и наноматериалов".

СПИСОК ЛИТЕРАТУРЫ

1. *Arnheim N.* Polymerase Chain Reaction Strategy // Annual review of biochemistry. 1992. V. 61. P. 131–156.
2. *Erlich H.A., Gelfand D., Sninsky J.J.* Recent Advances in the Polymerase Chain Reaction // Sciences. 1991. V. 252, N 5013. P. 1643–1651.
3. *Higuchi R., Fockler C., Dollinger C., Watson R.* Kinetic PCR Analysis: Real-Time Monitoring of DNA Amplification Reactions // Biotechnology. N. Y., Sept. 1993. V. 11, N 9. P. 1026–1030.
4. *Mullis K.B., Faloona F.A.* Specific Synthesis of DNA in Vitro via a Polymerase-Catalyzed Chain Reaction // Method in Enzymology. 1987. V. 155. P. 335–350.
5. *White T.J., Madej R., Persing D.H.* The Polymerase Chain Reaction: Clinical Applications // Advances in Clinical Chemistry. 1992. V. 29. P. 161–196.
6. *Livak K.J., Flood S.A., Marmaro J. et al.* Oligonucleotides with Fluorescent Dyes at Opposite Ends Provide a Quenched Probe System Useful for Detecting PCR Product and Nucleic Acid Hybridization // PCR Methods and Applic. 1995. N 4. P. 357–362.

7. Чернышев А.В., Друца В.Л. Проблемы создания оборудования для медицинской ПЦР-диагностики // Биомедицинские технологии и радиоэлектроника. 2004. № 12. С. 18.
8. Чернышев А.В. Создание теории рабочих процессов, методов расчета и разработка оборудования для ПЦР-диагностики. Автореф. дисс. ... доктора техн. наук. М.: МГТУ (МВТУ им. Н.Э. Баумана), 2006. 38 с.
9. Камке Э. Справочник по обыкновенным дифференциальным уравнениям. Пер. с нем. М.: Наука, 1976. С. 38–39.
10. Архипов Д.Б., Буляница А.Л., Щербаков А.П. Моделирование тенденций аналитического приборостроения логистическим уравнением с коэффициентами, зависящими от времени // Сб. "Тенденции развития аналитического приборостроения". СПб.: Русская классика, 2008. С. 20–21.
11. Горяинов В.Т., Журавлев А.Г., Тихонов В.И. Статистическая радиотехника: Примеры и задачи. Учеб. пособие для вузов. М.: Сов. радио, 1980. С. 50.
12. Первозванский А.А. Курс теории автоматического управления. Учеб. пособие. М.: Наука, 1986. 616 с.
13. Буляница А.Л., Курочкин В.Е. Исследование свойств и программно-аппаратная реализация алгоритма стохастической аппроксимации в модификации Я.З. Цыпкина // Научное приборостроение. 2002. Т. 12, № 2. С. 30–49.
14. Буляница А.Л., Курочкин В.Е., Бурьлов Д.А. Реализация процедуры оценивания постоянного сигнала на основе метода стохастической аппроксимации в модификации Я.З. Цыпкина // Радиотехника и электроника. 2002. Т. 47, № 3. С. 343–346.

*Институт аналитического приборостроения РАН,
Санкт-Петербург*

Материал поступил в редакцию 24.04.2009.

METHODS FOR LOGISTIC GROWTH CURVE PARAMETERS ESTIMATION. PART. 1. OPTIMIZATION OF ESTIMATION CONDITIONS AT THE PRESENCE OF ADDITIVE RANDOM ERROR

A. L. Bulyanitsa

Institute for Analytical Instrumentation RAS, Saint-Petersburg

Logistic growth is one of the traditional mathematical models, used for describing many population processes in particular, the dynamics of DNA quantity in polymerase chain reaction in real time (PCR-RT) is adequately described by a similar model. The parameters of the logistic growth curve serve the basis for the development of the necessary analytical information about the initial quantity of genetic material. The subject of the study is the sequence of measurements, each of which, in a general case, besides useful informative component contains random centered additive error and systematic component as a linear trend of the first order with a priori unknown parameters. The aim of the present part of the study is the development of the algorithms of the effective parameters estimation of logistic growth in the presence of the additive centered limited random error.

Keywords: polymerase chain reaction (PCR), logistic growth, estimation of parameters, random error, stochastic approximation