

КУМУЛЯНТНЫЙ АНАЛИЗ ЧИСЛА ФОТООТСЧЕТОВ ОТ ДВУХКОМПОНЕНТНЫХ СИСТЕМ ВО ФЛУОРЕСЦЕНТНОЙ ФЛУКТУАЦИОННОЙ СПЕКТРОСКОПИИ

А. Л. Науменок

Методы флуоресцентной флукуационной спектроскопии (ФФС) с каждым годом находят всё более широкое применение в химии, биофизике и медицине. Причиной этого служит присущая этим методам высокая чувствительность и неинвазивность, что делает их особенно удобными при исследовании живых клеток [1]. Одним из методов ФФС является кумулянтный анализ числа фотоотсчётов, позволяющий оценивать концентрацию и излучательные свойства флуоресцирующих молекул. В данной работе рассматривается подход к решению основной системы кумулянтного анализа для двухкомпонентного случая путём сведения ее к уравнению n -й степени.

В ходе эксперимента в ФФС с помощью конфокального микроскопа в образце создаётся малая область, в которой возбуждаются и излучают флуоресцирующие молекулы. На излучение молекул оказывают влияние различные процессы, в которых они участвуют, а также внутреннее строение самих молекул. Для аппроксимации распределения интенсивности регистрируемой флуоресценции молекул часто используется функция вида [2]

$$B(\mathbf{r}) = B_0 \exp \left[-\frac{2(x^2 + y^2)}{\omega_0^2} - \frac{2z^2}{z_0^2} \right], \quad (1)$$

где ω_0 и z_0 – заданные параметры.

Кумулянтный анализ (Fluorescence Cumulant Analysis – FCA) позволяет различать вещества по характеристической яркости молекул [3]. Система уравнений FCA, связывающая факториальные кумулянты K распределения числа фотоотсчётов (РЧФ) с оцениваемыми параметрами, имеет вид

$$K_i = \sum_{k=1}^M (q_k T)^i \gamma_i N_{effk}, \quad (2)$$

где $N_{effk} = c_k N_A V_{eff}$ – число частиц k -го типа в эффективном объёме $V_{eff} = \pi^{3/2} \omega_0^2 z_0$, q_k и c_k – характеристическая яркость и концентрация молекул k -го типа, $\gamma_i = \int B^i(\mathbf{r}) d\mathbf{r} / \int B(\mathbf{r}) d\mathbf{r}$, N_A – число Авогадро T – время подсчёта фотонов, используемое в РЧФ [4], на основе которого вычислены кумулянты K , $i=1, 2, \dots$

Для определения q_k и N_{effk} требуется вычислять статистически достоверные (относительная погрешность < 1) кумулянты. Чем выше порядок

кумулянта, тем больше погрешность его вычисления, поэтому на практике редко используются кумулянты выше 5-го порядка [3]. Анализ экспериментальных кумулянтов основан на подгонке к ним теоретически рассчитанных кумулянтов, зависящих от заданных параметров, с использованием некоторого критерия согласия.

Часто бывает известно отношение характеристических яркостей и (или) концентраций (количества частиц в эффективном объёме). Этой дополнительной информацией удобно пользоваться для повышения точности оценок параметров, найденных из системы (2).

Система уравнений может быть решена численно, однако в этом случае приходится подбирать начальные приближения для параметров q и N_{eff} . Одним из подходов решения системы (2) является ее сведение к уравнению n -й степени, корни которого находятся численно. Это позволяет снизить вычислительные затраты и использовать более простые алгоритмы для нахождения оценок параметров.

В общем случае используются все четыре уравнения системы:

$$A_i = N_1 q_1^i + N_2 q_2^i \quad (3),$$

где $A_i = K_i / \gamma_2 \gamma_i T^i$, $i=1,2,3,4$.

Система (3) сводится к уравнению 3-й степени вида:

$$H_1 q_2^3 + H_2 q_2^2 + H_3 q_2 + H_4 = 0, \quad (4)$$

где $H_1 = A_1(A_1 A_3 - A_2^2)$, $H_2 = A_3^2 - A_1^2 A_4$, $H_3 = -A_1 A_3^2 - A_2^2 A_3 + A_1 A_2 A_4$, $H_4 = A_2^2 A_4 + A_3^2 A_2$.

Получив корни полинома, можно найти оценки остальных параметров:

$$N_2 = \frac{A_2^2 - A_1 A_3}{2 A_2 q_2^2 - q_2^3 A_1 + A_3 q_2}, \quad N_1 = \frac{(A_1 - N_2 q_2)^2}{A_2 - q_2^2 N_2}, \quad q_1 = \frac{A_1 - N_2 q_2}{N_1}.$$

В случае известного соотношения количества частиц $N_2 = \Lambda$ для нахождения оценок параметров достаточно трёх уравнений:

$$A_i = N_1 (q_1^i + q_2^i \Lambda) \quad (5)$$

В ходе сведения данной системы к полиному возникает квадратное уравнение относительно q_2 , поэтому требуется рассматривать два набора корней этого уравнения. Это приводит к двум различным уравнениям 6-й степени вида:

$$\sum_{k=0}^6 h_k q_1^k = 0, \quad (6)$$

где h_k зависят от α и β . В итоге получается 12 наборов различных корней.

В случае известного соотношения молекулярных яркостей $q_2 = q_1 \alpha$ для решения достаточно трёх уравнений

$$A_i = q_1^i (N_1 + N_2 \alpha^i). \quad (7)$$

Система уравнений сводится к квадратному уравнению с известными коэффициентами

$$a q_1^2 + b q_1 + c = 0, \quad (8)$$

где $a = A_1 \alpha (1 - \alpha)$, $b = A_2 (\alpha^2 - 1)$, $c = A_3 (1 - \alpha)$.

В случае с известными соотношениями характеристических яркостей $q_2 = q_1 \alpha$ и количества частиц $N_2 = N_1 \beta$ для решения достаточно двух уравнений, при этом повышается точность вычислений, так как не нужно использовать дополнительные кумулянты:

$$A_i = N_1 q_1^i (1 + \alpha^i \beta). \quad (9)$$

Система решается аналитически, в итоге получим

$$N_1 = \frac{A_1^2 (1 + \alpha^2 \beta)}{A_2 (1 + \alpha \beta)^2}, \quad q_1 = \frac{A_2 (1 + \alpha \beta)}{A_1 (1 + \alpha^2 \beta)}. \quad (10)$$

Параметры N_2 и q_2 находятся из известных соотношений.

На оценки параметров, полученных из наборов решений уравнений с полиномом не должны быть отрицательными или комплексными. Отбор наилучших оценок параметров производится с помощью некоторого критерия согласия, например, наилучшими являются те оценки параметров, которые доставляют минимум выражению $\sum (K_{ii} - K_i)^2$, где K_{ii} – кумулянты, рассчитанные из функции распределения фотоотсчётов, K_i – кумулянты, рассчитанные из набора корней.

При тестировании на незашумлённых данных в общем и всех частных случаях системы уравнения кумулянтного анализа для двухкомпонентного найденное численное решение имело погрешность меньше 1%. В ходе тестирования на зашумлённых данных, полученных с помощью имитационных моделей [5], было установлено, что общий случай и случай с известными характеристическими яркостями наиболее чувствительны к статистическому шуму. Наилучшие оценки параметров получаются при использовании с известными соотношениями характеристических яркостей и количества частиц.

Разработанный подход к решению основной системы кумулянтного анализа для двухкомпонентного случая позволяет сократить количество

используемых кумулянтов, что приводит к повышению точности оценок параметров. Нахождение корней полинома позволяет сократить время нахождения оценок параметров по сравнению со временем решения системы уравнений с помощью численных методов.

Литература

1. *Lakowicz J. R.* Principles of Fluorescence Spectroscopy. Singapore, 2006.
2. *Krichevsky O., Bonnet G.* Fluorescence correlation spectroscopy: the technique and its applications // Reports on Progress in Physics. 2002, V. 65, P. 251-297.
3. *Muller J. D.* Cumulant Analysis in Fluorescence Fluctuation Spectroscopy // Biophys. J. 2004, V. 86, P. 3981-3992.
4. *Chen Y., Müller J., So P., Gratton E.* The Photon Counting Histogram in Fluorescence Fluctuation Spectroscopy // Biophys. J. 1999, V. 77, P. 553-567.
5. *Shingaryov I., Skakun V., Apanasovich V.* Photon Counts Simulation in Fluorescence Fluctuation Spectroscopy // Pattern Recognition and Information Processing (PRIP), 2009, Minsk. P. 178-182.

АНАЛИЗ БИОЛОГИЧЕСКИХ ДНК-МИКРОЧИПОВ С ИСПОЛЬЗОВАНИЕМ СРЕДЫ R

А. С. Рындин

1. ВВЕДЕНИЕ

Биочипы, или микроматрицы ДНК, оказали огромное влияние на развитие медико-биологических дисциплин, связанных с исследованием генов, включая онкологию, токсикологию, фармакологию, биологию развития [1]. Эксперименты с участием биочипов позволяют изучать функции генов, их взаимосвязь, биологические процессы с их участием, а также проводить множество других биологических исследований [1, 2].

Эксперимент с биочипами выдаёт данные об экспрессии десятков тысяч генов. Проанализировать такой огромный объём данных можно только с помощью вычислительной мощности компьютеров. Подобный анализ огромных массивов биологических данных составляет предмет самостоятельной области науки – биоинформатики, научной дисциплины на стыке биологии, информатики и вычислительной математики.

Целью данной работы является изучение статистических методов обработки экспериментальных биологических данных, полученных с помощью биочипов, и их реализация в среде R на примере опубликованных данных [3].