

Статистический тест для выявления присутствия скрытых пиков в спектральных сигналах

К. К. Семенов

*Высшая школа компьютерных технологий и информационных систем
Санкт-Петербургского политехнического университета Петра Великого*

semenov_kk@spbstu.ru

Аннотация. В статье рассмотрена задача обнаружения скрытых пиков в спектральных сигналах – компонент, ставших незаметными вследствие свертки истинного спектра с аппаратной функцией спектроскопического средства измерений. Предложен статистический критерий, основанный на обобщенном тесте отношения правдоподобия, указано распределение используемой статистики и формулы для р-значения. Для снижения вероятностей ошибок первого и второго родов критерий дополнен рядом вспомогательных процедур: алгоритмом выделения спектральных участков с потенциальными скрытыми пиками; оценкой погрешности значения статистики, унаследованной ей от шумов измеренного спектра; робастной деконволюцией.

Ключевые слова: спектральные сигналы, аппаратная функция, спектроскопические средства измерений, скрытые пики, статистические гипотезы, критерии проверки статистических гипотез

I. ВВЕДЕНИЕ

Измерительный сигнал, получаемый на выходе любого спектроскопического средства измерений, представляет собой свертку истинного спектра с аппаратной функцией измерительного устройства. Записанный спектральный сигнал $y(\lambda)$ математически описывается как:

$$y(\lambda) = x(\lambda) * h(\lambda) + \varepsilon(\lambda),$$

где $x(\lambda)$ – действительный спектр, $h(\lambda)$ – аппаратная функция средства измерения, $\varepsilon(\lambda)$ – случайный шум, а знак «*» означает операцию свертки. Конечная ширина функции $h(\lambda)$ ограничивает способность средства измерений различать близко расположенные спектральные линии. В хроматографии аналогичный эффект обусловлен дисперсионными механизмами внутри колонки. Данное обстоятельство порождает задачу обнаружения скрытых пиков в спектре, «исчезнувших» при размытии спектра при свертке с аппаратной функцией прибора, являющейся одной из ключевых задач так называемой теории статистического перекрытия пиков (Statistical Overlap Theory, SOT) [1]: число видимых пиков систематически занижает число компонентов, присутствующих в анализируемом образце, причем при снижении эффективности разделения это занижение нарастает монотонно [2, 3].

Настоящая работа формулирует данную задачу в терминах теории проверки статистических гипотез: дан измеренный сигнал на заданном участке спектра, предположительно содержащий один или несколько перекрывающихся пиков; требуется на основании

статистического критерия сделать вывод о числе компонент.

II. ПОДХОДЫ К ВЫЯВЛЕНИЮ СКРЫТЫХ ПИКОВ В СПЕКТРАЛЬНЫХ СИГНАЛАХ

Подобный статистический подход к данной задаче либо близким к ним восходит к теории обработки радиотехнических сигналов (радары, акустика, антенные решетки) и был изначально разработан применительно к задаче разрешения близких источников. В частности, в указанной области вводят так называемый статистический предел разрешения, определяемый как минимальное расстояние между параметрами двух источников, при котором возможно их корректное разделение при заданных пределах для вероятностей ложно-положительных и ложно-отрицательных срабатываний [4, 5]. В работах [6, 7] был представлен критерий для среднего значения базовой линии спектра: два источника считаются разрешенными, если вторая производная базовой линии спектра в средней точке между соответствующими им спектральными элементами отрицательна. В [4] был введен бинарный тест на принятие решения о присутствии одного либо двух источников по результатам анализа данных на основе обобщенного теста отношения правдоподобия (Generalized Likelihood Ratio Test, GLRT) [8]. В упомянутой работе показано, что этот тест асимптотически является равномерно наиболее мощным, что в совокупности с тем обстоятельством, что статистика критерия не привязана к конкретному алгоритму оценивания, делает его предпочтительным для спектроскопических и иных применений, в том числе и в настоящей работе. Наконец, третьим основным подходом является оценочный подход на основе границы Крамера-Рао, которая, как известно, задает нижнюю границу дисперсии несмещенных оценок параметров. Предложенный в [5] статистический критерий на его основе сформулирован следующим образом: два источника разрешимы, если расстояние между ними превышает стандартное отклонение оценки этого расстояния, вычисляемое через границу Крамера-Рао.

Применение GLRT для задачи обнаружения скрытого пика в спектрограмме требует следующего: параметрической модели формы пика (как правило, Гауссова или лоренцева, или их свертка – функция Фойгта); знания или оценки ширины аппаратной функции (Full Width at Half Maximum, FWHM); модели шума (обычно нормальное распределение) [9, 10]. При выполнении этих условий GLRT обеспечивает контролируемый уровень ошибки первого рода, что является принципиальным отличием от эвристических подходов, включая перечисленные выше.

Исследование выполнено за счет гранта Российского научного фонда № 25-21-20109, <https://rscf.ru/project/25-21-20109/>, и средств гранта Санкт-Петербургского научного фонда № 25-21-20109.

Задача проверки гипотезы о наличии скрытого пика может быть сведена в задаче применения так называемых информационных критериев (Informational Criterion, IC), широко применяемых при кластеризации с целью определения числа групп, устойчиво выделяемых в анализируемой совокупности векторов. Такая постановка переводит задачу из поставленной локально (проверка наличия скрытого пика на заданном участке спектра) в поставленную глобально (определить, сколько гауссовых или иных компонент с точки зрения выбранного информационного критерия оптимально описывают спектр целиком). Среди применяемых на практике IC следует выделять критерий Акаике (AIC) и Байеса (BIC) [11, 12], причем, как отмечается в литературе, последний лучше подходит для идентификации истинного числа компонент (AIC склонен к переоценке их числа). Суть задачи в таком случае сводится к определению такого числа и положения спектральных компонент, которые обеспечивают минимум значения информационного критерия. Принципиальное ограничение BIC в задаче обнаружения скрытого пика состоит в том, что применение BIC не является статистическим тестом в строгом смысле и не контролирует вероятность ошибки первого рода. Кроме того, смесевые модели склонны к переобучению, поэтому в литературе не рекомендуется применять критерии AIC/BIC для проверки гипотез о числе пиков [13].

Другими распространенными нестатистическими способами по обнаружению скрытых пиков являются методы, направленные на восстановление исходного спектра $x(\lambda)$, – в случае успешного выполнения данной операции незаметные в измеренном спектре $y(\lambda)$ пики становятся видимыми. Основными методами являются методы деконволюции на основе преобразования Фурье [14] и методы разложения хроматографического или спектрального пика на гауссовы (или иные) компоненты методом наименьших квадратов [15]. Данные подходы требуют априорного задания числа пиков и могут быть использованы для обнаружения скрытых пиков лишь итерационно и без контроля вероятностей принятия ошибочного решения. На практике также часто используют подходы на основе вычисления второй производной спектра $d^2y(\lambda)/d\lambda^2$ для выявления так называемых плеч в спектре и скрытых пиков, которым соответствуют минимумы значений $d^2y(\lambda)/d\lambda^2$, поскольку они не требуют, но в высшей степени чувствительны к присутствию шума и выбору метода оценки значения производной [16, 17].

Особо остро стоит задача выявления скрытых пиков для ядерной магнитно-резонансной спектроскопии биологических образцов, поскольку перекрытие соответствующих спектрографических сигналов наблюдается для большинства метаболитов [18]. Для данной области в последние годы характерны следующие подходы. Предложенный в [19] программный комплекс UnidecNMR идентифицирует резонансы в спектрах через итерационную деконволюцию для данных произвольной размерности, толерантен к неправильному выбору параметра формы пика и для сильно перекрытых спектров олигосахаридов и белков демонстрирует результативность, сопоставимую с экспертным ручным анализом. Работы [20, 21] описывают вероятностный подход к наличию пика в той или иной области анализируемого спектра (P3

– Peak Probability Presentation): каждой точке спектра присваивается вероятность того, что в ней расположен максимум пика. Соответствующие оценки осуществляются с применением физически-информированной нейронной сети (Physics-Informed Neural Network, PINN). Было показано [20], что точность локализации пиков с помощью такого подхода приближается к границе Крамера-Рао, то есть к теоретическому минимуму дисперсии несмещенных оценок. Данный подход представляет собой фактически задачу статистического вывода, в которой нейронная сеть заменяет явную параметрическую модель происходящих при измерении искажений спектра. Вместе с тем подход P3 не позволяет задать формальный уровень значимости без дополнительной калибровки PINN.

Подход на основе обобщенного теста отношения правдоподобия позволяет построить статистический критерий для проверки гипотезы о присутствии скрытого пика в спектральном сигнале. Для успешного использования GLRT на практике в его классической постановке необходимо задание модели формы пика и аппаратной функции средства измерений, а также модели шума. Далее в работе показано, каким образом может быть построен статистический тест, свободный от ряда из перечисленных ограничений.

III. СТАТИСТИЧЕСКИЙ ТЕСТ ОПРЕДЕЛЕНИЯ ДОСТАТОЧНЫХ ОСНОВАНИЙ ПОЛАГАТЬ ПРИСУТСТВИЕ СКРЫТОГО ПИКА

В данной работе был разработан статистический тест, направленный на выявление скрытых пиков в анализируемых сигналах измерительной информации спектрографического вида.

Пусть анализируемый участок, на котором определяется присутствие скрытого пика, размытого вследствие воздействия достаточно широкой аппаратной функции применяемого средства измерений, представлен отсчетами сигнала, выраженными в виде вектора значений $\mathbf{y}^T = (y_1, y_2, \dots, y_n)$. Номера отсчетов поставлены в соответствие с началом анализируемого участка. Значения, входящие в вектор \mathbf{y} могут быть описаны следующим образом: $\mathbf{y} = \mathbf{h} * (\mathbf{x} + \alpha \cdot \mathbf{g}) + \boldsymbol{\varepsilon}$, где \mathbf{h} – это вектор значений, описывающий аппаратную функцию средства измерений (должна быть либо известна или оценена), ответственную за размытие сигнала измерительной информации; \mathbf{x} – вектор, описывающий значения, описывающие основную компоненту анализируемого сигнала (например, доминирующий широкий пик или фон); \mathbf{g} – вектор, описывающий стандартную форму предполагаемого скрытого пика (нормированная гауссиана), положение μ и параметр масштаба σ (определяет ширину пика) которого неизвестны; $\alpha \geq 0$ – амплитуда (интенсивность) скрытого пика; $*$ – операция свертки, моделирующая размытие истинного спектра $(\mathbf{x} + \alpha \cdot \mathbf{g})$ аппаратной функцией \mathbf{h} ; $\boldsymbol{\varepsilon}^T = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ – вектор погрешностей и шумов, которые предполагаются независимыми, нормально распределенными и не имеющими систематической составляющей, то есть $\varepsilon_i \sim N(0, s^2)$. Дисперсия шума s^2 может быть оценена.

Выдвигаются две гипотезы.

H₀ (нулевая): скрытого пика нет $\Rightarrow \alpha = 0$.

H₁ (альтернативная): скрытый пик присутствует $\Rightarrow \alpha > 0$.

Соответственно задача проверки сформулированных утверждений сводится к проверке значимости параметра α на границе области его возможных значений. Как было отмечено, в качестве статистики теста предлагается выбрать обобщенное отношение правдоподобий, как позволяющее добиться наибольшей мощности при статистической проверке. Значение статистики определяется следующим набором шагов.

Вначале осуществляется построение функции правдоподобия. Поскольку шум является гауссовым, то условная плотность наблюдений при указанных параметрах (α, θ) имеет вид:

$$L(\alpha, \theta; \mathbf{y}) = \frac{\exp\left(-\frac{1}{2 \cdot \sigma^2} \cdot \left\| \mathbf{y} - (\mathbf{h} * (\mathbf{x} + \alpha \cdot g(\theta))) \right\|^2\right)}{(2 \cdot \pi \cdot \sigma^2)^{\frac{n}{2}}},$$

где $\theta^T = (\mu, \sigma)$ – вектор всех дополнительных параметров (например, положение скрытого пика, ширина, параметры фона).

Соответственно обобщенное отношение правдоподобий вычисляется как

$$GLR = \frac{\sup_{\theta, \alpha > 0} L(\alpha, \theta; \mathbf{y})}{\sup_{\theta} L(0, \theta; \mathbf{y})} = \frac{L_{\max}(\mathbf{H}_1)}{L_{\max}(\mathbf{H}_0)},$$

где $L_{\max}(\mathbf{H}_1)$ – максимум правдоподобия при альтернативной гипотезе (скрытый пик есть), $L_{\max}(\mathbf{H}_0)$ – максимум правдоподобия при нулевой гипотезе (скрытого пика нет).

В качестве статистики предлагается использовать значение логарифма от GLR^2 :

$$R = 2 \cdot \ln GLR = 2 \cdot (\ln L_{\max}(\mathbf{H}_1) - \ln L_{\max}(\mathbf{H}_0)).$$

Выполнив максимизацию функционала $L(\alpha, \theta; \mathbf{y})$ по θ и по α (с ограничением $\alpha > 0$), статистика R может быть упрощена до следующего вида:

$$R = \max \left\{ 0, \frac{((\mathbf{y} - (\mathbf{h} * \mathbf{x})) \times (\mathbf{h} * \mathbf{g}^*))^2)}{s^2 \cdot \|\mathbf{h} * \mathbf{g}^*\|^2} \right\}$$

где $(\mathbf{y} - (\mathbf{h} * \mathbf{x}))$ – вектор невязок после подгонки модели анализируемого участка сигнала только под основную компоненту (т.е. при $\alpha=0$); $\mathbf{h} * \mathbf{g}^*$ – вектор, описывающий модель скрытого пика (после воздействия на него аппаратной функции); \mathbf{g}^* – оценка формы скрытого пика (например, определяется перебором возможных вариантов); \times – операция скалярного произведения двух векторов; $\|\mathbf{h} * \mathbf{g}^*\|^2 = (\mathbf{h} * \mathbf{g}^*)^T \cdot (\mathbf{h} * \mathbf{g}^*)$ – норма вектора $\mathbf{h} * \mathbf{g}^*$. Сравнение с нулем выполняется по той причине, что пик по умолчанию предполагает положительное значение амплитуды. Таким образом, статистика R по сути дела показывает, насколько хорошо модель скрытого пика, подвергнувшегося размытию вместе с основным сигналом, объясняет структуру остатков $(\mathbf{y} - (\mathbf{h} * \mathbf{x}))$ по сравнению с уровнем шума s^2 .

Значения данной статистики имеют следующее распределение. Из вида статистики R было бы естественным предположить, что ее асимптотическое распределение совпадает с законом «хи-квадрат», однако в силу того, что проверка амплитуды α осуществляется на границе области ее возможных значений ($\alpha \geq 0$) и часть параметров θ может быть не идентифицирована при

нулевой гипотезе (например, положение скрытого пика невозможно определить, если его амплитуда равна нулю), то в силу работы [22] распределение статистики R при нулевой гипотезе \mathbf{H}_0 является смесью распределений хи-квадрат с нулем и одной степенями свободы соответственно:

$$R \in (0, 5 \cdot \delta_0 + 0, 5 \cdot \chi^2(1)),$$

где $\delta_0 = \begin{cases} +\infty, & x = 0, \\ 0, & x \neq 0 \end{cases}$ – дельта-функция Дирака в нуле, $\int_{-A}^A \delta_0(x) dx = 1$ при любых $A > 0$. Соответственно критическое значение R_γ для уровня значимости γ (например, $\gamma = 0,05$) для статистики R следует находить из уравнения

$$P(R > R_\gamma | \mathbf{H}_0) = 0,5 \cdot P(\chi^2(1) > R_\gamma) = \gamma.$$

Следовательно, $R_\gamma = \chi_{1-2 \cdot \gamma}^2(1)$, где $\chi_{1-2 \cdot \gamma}^2(1)$ – $(1 - 2 \cdot \gamma) \cdot 100\%$ -ая квантиль распределения «хи-квадрат» с одной степенью свободы.

P -значение (p -value) для рассматриваемого статистического теста следует определять соответственно по соотношению:

$$p = 0,5 \cdot P(\chi^2(1) > R) = 0,5 \cdot (1 - P(\chi^2(1) < R)) = 0,5 \cdot (1 - F_{\chi^2(1)}(R)),$$

где $F_{\chi^2(1)}(R)$ – функция распределения (cdf) для случайной величины, имеющей плотность распределения $\chi^2(1)$. Так как $F_{\chi^2(1)}(R) = \frac{\hat{\Gamma}(1/2, R/2)}{\Gamma(1/2)} = \frac{\hat{\Gamma}(1/2, R/2)}{\sqrt{\pi}}$, то

$$p = 0,5 \cdot \left(1 - \frac{1}{\sqrt{\pi}} \cdot \hat{\Gamma}\left(\frac{1}{2}, \frac{R}{2}\right)\right) = 0,5 \cdot \left(1 - \frac{1}{\sqrt{\pi}} \cdot \int_0^{\frac{R}{2}} t^{-\frac{1}{2}} \cdot e^{-t} dt\right).$$

Здесь $\Gamma(z)$ – гамма-функция, $\hat{\Gamma}(z_1, z_2)$ – неполная гамма-функция.

Безусловно, при большом количестве перебираемых вариантов местоположения скрытого пика и его потенциальной ширины в тест необходимо включать поправку Бонферрони на множественные проверки: скорректированный уровень значимости γ_{corr} оказывается равен $\gamma_{corr} = \gamma/m$, где m – число совершаемых проверок на одном и том же участке анализируемого сигнала для разных возможных параметров скрытого пика. Соответственно, для каждого отдельного производимого теста следует использовать критическое значение

$$R_{\gamma_{corr}} = \chi_{1-2 \cdot \gamma_{corr}}^2 \quad (1)$$

или, что эквивалентно, сравнивать p -value не со значением $\gamma = 0,05$, а со значением $\gamma_{corr} = \gamma/m$ при принятии решения о принятии или отклонении нулевой гипотезы в каждом отдельном тестировании.

IV. УМЕНЬШЕНИЕ ВЕРОЯТНОСТЕЙ ОШИБОК 1 и 2 РОДА

Для уменьшения вероятностей неверного принятия решений представленный статистический критерий был дополнен следующими процедурами.

1) Был предложен алгоритм выделения фрагментов спектра, содержащих потенциальные участки со скрытыми пиками: в качестве таковых было решено принимать участки между соседними достоверно

локализуемыми пиками в сигнале измерительной информации. Обнаружение локальных максимумов проиллюстрировано на рис. 1 и 2 ниже. Для снижения числа ложных срабатываний эвристически определяется минимально возможная ширина пика, завершающегося локальным максимумом (обычно исходя из разрешающей способности используемого средства измерений – наблюдаемый пик не должен иметь ширину длиной ниже предела чувствительности).

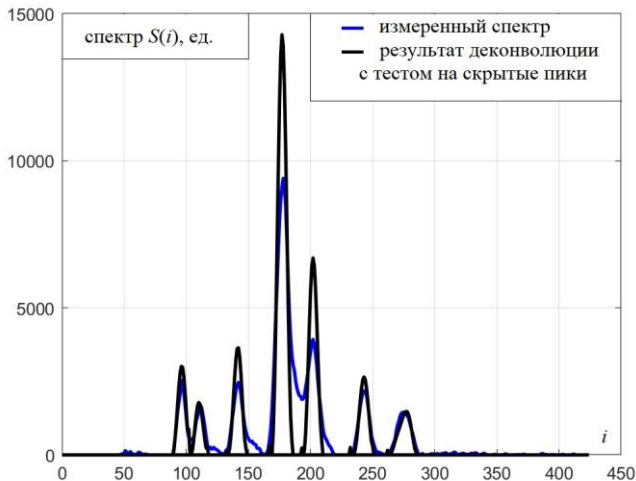


Рис. 1. Пример применения теста на присутствие скрытого пика для спектра с явно выраженными пиками: $p\text{-value} = 0,99$

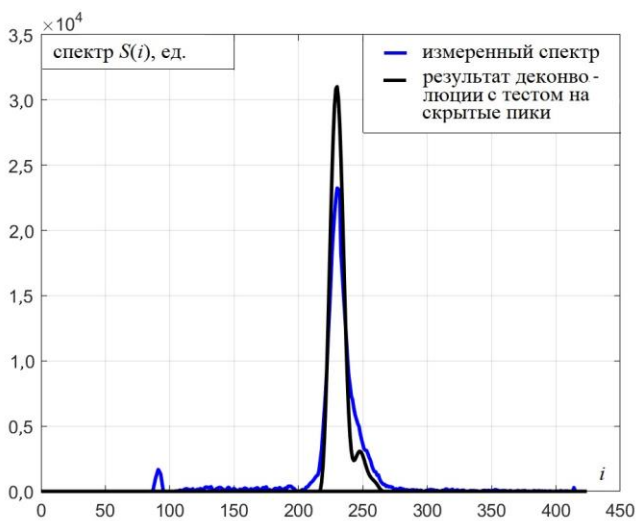


Рис. 2. Пример применения теста на присутствие скрытого пика для спектра с неявно выраженными пиками: $p\text{-value} = 0,10$

2) Был использован алгоритм метрологического сопровождения вычислений с неточными данными [23], необходимый для автоматической оценки параметра регуляризации при решении задачи деконволюции. Данный подход упрощенно заключается в следующем. Пусть $t = f(\mathbf{y})$ – результат вычислений в соответствии с некоторой функцией f , зависящей от вектора аргументов $\mathbf{y}^T = (y_1, y_2, \dots, y_n)$, значение каждого из которых известно с погрешностью Δy_i , $i = 1, 2, \dots, n$, не превосходящей по модулю соответственно предельных значений $\Delta^T = (\Delta_1, \Delta_2, \dots, \Delta_n)$. Тогда в расчете на наихудший случай предел возможной погрешности величины t не превзойдет в первом приближении величины

$$\Delta_t \approx \sum_{i=1}^n \left| \frac{\partial f(\mathbf{y})}{\partial y_i} \right| \cdot \Delta_i,$$

а в случае, если у всех компонент вектора \mathbf{y} погрешности не зависят друг от друга и всем пределам погрешности Δ_i соответствует одно и то же значение доверительной вероятности $Q = 0,95$,

$$\Delta_t \approx \sqrt{\sum_{i=1}^n \left(\frac{\partial f(\mathbf{y})}{\partial y_i} \right)^2 \cdot \Delta_i^2}.$$

Чтобы оценить указанные значения, необходимо определять значения частных производных $\partial f(\mathbf{y})/\partial y_i$, что можно сделать тем или иным способом – начиная от проблемного метода конечных разностей и заканчивая перспективным методом автоматического дифференцирования [24]. Также значение величины Δ_t может быть получено стохастическим моделированием, когда для входных аргументов \mathbf{y} случайным образом определяются их возможные действительные значения, извлекаемые из многомерного бруса $(\mathbf{y}-\Delta, \mathbf{y}+\Delta)$.

Применение данного подхода позволяет после оценки величины s возможного среднеквадратического отклонения случайных погрешностей каждого отдельного значения анализируемого участка \mathbf{y} определить величину возможной случайной погрешности статистики R , унаследованной ей от анализируемого сигнала. Получение соответствующего предела погрешности Δ_R позволяет более обоснованно принять решение при проверке гипотезы о присутствии / отсутствии скрытого пика на анализируемой участке сигнала измерительной информации:

- если $R - \Delta_R > R_Y$, то следует отклонить нулевую гипотезу;
- если $R + \Delta_R < R_Y$, то следует отклонить альтернативную гипотезу;
- в случае, если $R_Y \in [R - \Delta_R, R + \Delta_R]$ ситуация не определена и принимать решение следует, исходя из потенциальных рисков неверного решения.

3) Был использован алгоритм робастной деконволюции [25] на основе принципа минимального модуля, произвольных моделей аппаратной функции используемого спектроскопического прибора (с функцией Фойгта в качестве модели по умолчанию) и метода Говарда [26] для восстановления неотрицательных сигналов для удаления эффекта Гиббса.

Данный подход, реализующий операцию деконволюции $*^{-1}$, в сочетании с изложенным в п.2 данного перечня подходом к оценке величины погрешности, наследуемой результатом вычислений от исходных данных, позволяет структурно (но не вычислительно) упростить рассматриваемый критерий проверки гипотезы о скрытом пике и сделать его менее параметрическим. Пусть в составе анализируемого участка сигнала \mathbf{y} в заданной позиции i проверяется наличие скрытого пика. Для заданной модели аппаратной функции средства измерения, с помощью которого получен сигнал \mathbf{y} , осуществляется деконволюция $(\mathbf{y} *^{-1} \mathbf{h})$, причем ширина аппаратной функции \mathbf{h} подбирается так, чтобы в точке i оказался выделен отдельный пик \mathbf{x} , не пересекающийся с соседними пиками. Если это оказывается невозможным

(данный участок сигнала не отделяется от соседних пиков при всех значениях ширины аппаратной функции в заданных физических пределах), то пик однозначно не обнаруживается. Если удастся пик отделить, то определяются его амплитуда α и величина предела его среднеквадратической погрешности σ_α , унаследованной от погрешности y (случайного шума со среднеквадратическим отклонением s). В качестве статистики выбирается стандартный z-score: $z = \alpha/\sigma_\alpha$. Если $z < N_{1-\gamma}(0,1)$, то следует принять гипотезу H_1 , если же $z > N_{1-\gamma}(0,1)$, то следует принять H_0 . Здесь $N_{1-\gamma}(0,1)$ – $(1-\gamma)\cdot 100\%$ -ая квантиль стандартного нормального закона распределения (т.е. с нулевым математическим ожиданием и единичной дисперсией).

Представленный критерий обладает улучшенными статистическими свойствами – при этом, и для него обнаружение скрытого пика, полностью поглощенного «крылом» доминирующего пика без какого-либо видимого плеча (рис. 2), остается задачей на пределе возможностей и определяется прежде всего фундаментальными информационными ограничениями в виде границы Крамера-Рао.

V. ЗАКЛЮЧЕНИЕ

В данной работе представлен статистический критерий, основанный на обобщенном тесте отношения правдоподобия, для проверки гипотезы о присутствии на анализируемом участке спектрального сигнала скрытого пика. Представленный критерий дополнен специальными процедурами, позволяющими снизить вероятности ошибок первого и второго родов при принятии решения, включая метод робастной деконволюции для учета всей полноты информации, содержащейся в анализируемом участке спектра.

СПИСОК ЛИТЕРАТУРЫ

- [1] Davis, J.M. (1997). Extension of statistical overlap theory to poorly resolved separations. *Analytical Chemistry*, 69(18), 3796-3805.
- [2] Davis, J.M., Giddings, J.C. (1983). Statistical theory of component overlap in multicomponent chromatograms. *Analytical Chemistry*, 55(3), 418-424.
- [3] Schure, M.R., Davis, J.M. (2015). The simple use of statistical overlap theory in chromatography. *LCGC North America*, 33, 14-18.
- [4] El Korso, M.N., Boyer, R., Renaux, A., Marcos, S. (2011). Statistical resolution limit for the multidimensional harmonic retrieval model: hypothesis test and Cramér-Rao Bound approaches. *EURASIP Journal on Advances in Signal Processing*, 2011(1), 12.
- [5] Smith, S.T. (2005). Statistical resolution limits and the complexified Cramer-Rao bound. *IEEE Transactions on Signal Processing*, 53(5), 1597-1609.
- [6] Cox, H. (1973). Resolving power and sensitivity to mismatch of optimum array processors. *The Journal of the Acoustical Society of America*, 54(3), 771-785.
- [7] Sharman, K., Durrani, T. (1985). Resolving power of signal subspace methods for finite data lengths. In *ICASSP'85. IEEE International Conference on Acoustics, Speech, and Signal Processing* (vol. 10, pp. 1501-1504). IEEE.
- [8] Zeitouni, O., Ziv, J., Merhav, N. (1992). When is the generalized likelihood ratio test optimal? *IEEE Transactions on Information Theory*, 38(5), 1597-1602.
- [9] Fan, J., Jiang, J. (2007). Nonparametric inference with generalized likelihood ratio tests. *Test*, 16(3), 409-444.
- [10] Cavassila, S., Deval, S., Huegen, C., van Ormondt, D., Graveron-Demilly, D. (2000). Cramer-Rao bound expressions for parametric estimation of overlapping peaks: influence of prior knowledge. *Journal of Magnetic Resonance*, 143(2), 311-320.
- [11] Zanini, A., Woodbury, A.D. (2016). Contaminant source reconstruction by empirical Bayes and Akaike's Bayesian Information Criterion. *Journal of contaminant hydrology*, 185, 74-86.
- [12] Granlund, L., Billinge, S.J.L., Duxbury, P.M. (2015). Algorithm for systematic peak extraction from atomic pair distribution functions. *Foundations of Crystallography*, 71(4), 392-409.
- [13] Tiley, G.P., Barker, M.S., Burleigh, J.G. (2018). Assessing the performance of Ks plots for detecting ancient whole genome duplications. *Genome biology and evolution*, 10(11), 2882-2898.
- [14] Jansson, P.A. (Ed.). (2014). *Deconvolution of images and spectra*. Courier Corporation. 528 p.
- [15] Bourguignon, S., Mary, D., Slezak, É. (2011). Restoration of astrophysical spectra with sparsity constraints: Models and algorithms. *IEEE Journal of Selected Topics in Signal Processing*, 5(5), 1002-1013.
- [16] Hegland, M., Anderssen, R.S. (2005). Resolution enhancement of spectra using differentiation. *Inverse Problems*, 21(3), 915-934.
- [17] Merrick, M.F., Pardue, H.L. (1986). Evaluation of absorption and first- and second-derivative spectra for simultaneous quantification of bilirubin and hemoglobin. *Clinical chemistry*, 32(4), 598-602.
- [18] Sharma, R., Gogna, N., Singh, H., Dorai, K. (2017). Fast profiling of metabolite mixtures using chemometric analysis of a speeded-up 2D heteronuclear correlation NMR experiment. *Rsc Advances*, 7(47), 29860-29870.
- [19] Buchanan, C., Karunanithy, G., Tkachenko, O., et al. (2025). UnidecNMR: automatic peak detection for NMR spectra in 1-4 dimensions. *Nature Communications*, 16(1), 449.
- [20] Jahangiri, A., Agback, T., Brath, U., Orekhov, V. (2026). Toward ultimate NMR resolution with deep learning. *Science Advances*, 12(13), eady7995.
- [21] Zhan, H., Huang, Y., Chen, Z. (2026). Recent progress in artificial intelligence enabled NMR spectroscopy: Methodologies, implementations, quality assessments, and prospects. *Applied Physics Reviews*, 13(1), 011322.
- [22] Self, S.G., Liang, K.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398), 605-610.
- [23] Семенов, К.К. Программные средства обработки неточных данных, выполняемой в средствах измерений // *Известия Самарского научного центра Российской академии наук*. 2016. Т. 18(4-7). С. 1444-1455.
- [24] Семенов, К.К. Автоматическое дифференцирование функций, выраженных программным кодом // *Известия высших учебных заведений. Приборостроение*. 2011. Т. 54(12). С. 34-40.
- [25] Semenov, K., Taraskin, A., Yurchenko, A., Baranovskaya, I., Purvinsh, L., Gyulikhandanova, N., Vasin, A. (2023). Uncertainty estimation for quantitative agarose gel electrophoresis of nucleic acids. *Sensors*, 23(4), 1999.
- [26] Howard, S.J. (1981). Continuation of discrete Fourier spectra using a minimum-negativity constraint. *J. Opt. Soc. Am.*, 71, 819-824.