

ХАРАКТЕРНЫЕ ОСОБЕННОСТИ СТАТИСТИЧЕСКОГО ПОДХОДА В СОВРЕМЕННЫХ БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

А. А. ОРДУХАНИЯН, Р. М. АРУТЮНЯН, Э. А. ДАНИЕЛЯН

РНБЦ МЗ Армянской ССР, Ереван, Ереванский государственный университет

Аннотация — Обсуждаются характерные особенности статистического подхода в биологии на современном этапе. Предложен сценарий биоматематического моделирования в качестве адекватного подхода при решении конкретных биологических задач. Универсальность подхода продемонстрирована на примерах его использования авторами в различных областях (фармакология, иммунология, физиология, цитогенетика).

Ստատիստիկա — Բնութագրված են կենսաբանության մեջ ժամանակակից փուլի գիտական հարստացման մոտեցման բնորոշ յուրահատկությունները: Առաջարկված է կենսամաթեմատիկական մոդելավորման սցենար, որը հանդիսանում է ազնվատ մոտեցում կոնկրետ կենսաբանական խնդիրների լուծման համար: Մոտեցման բազմակողմանեությունը դաճագրված է տարբեր բնագավառներում ճշգրիտների փորձից բխող օրինակները փրկ չնարմակողոզիս, իմունոլոզիս, ֆիզիոլոզիս, սիտոգենեթիկա):

Abstract — The characteristic aspects of statistical approach in modern biological application are discussed. The scenario of biostatistical modelling is proposed. It shows its adequacy for the solving of concrete biological problems. The universality of the proposed approach is demonstrated on the examples, that are the result of authors own experience in various branches (pharmacology, immunology, physiology, cytogenetics).

Ключевые слова: моделирование биоматематическое, многокритерный статистический анализ, функциональные системы организма.

В моделировании биологических процессов выделяются два направления: детерминистическое и стохастическое (статистическое). Детерминистическое описание биологических объектов в условиях неопределенности обычно затруднено. В биомедицинских исследованиях неполнота информации о ряде процессов диктует необходимость при их моделировании использовать статистический подход.

Статистический подход в собственно биологии неоднократно обсуждался, мы же будем концентрировать внимание на накопленном опыте использования биоматематического моделирования (БММ) в интерпретации Пайла [20]. Обсуждение статистического подхода и моделирования на современном этапе позволяет выделить ряд особенностей и рекомендовать иллюстрированную примерами программу использова-

ния БММ в конкретных прикладных биологических исследованиях. На наш взгляд, при этом следует руководствоваться следующими положениями.

1. *Многомерность БММ*—отражает сложность биологических объектов, следствием чего является взаимосвязь параметров соответствующей математической модели; это диктует необходимость привлечения многомерного статистического анализа с элементами теории распознавания образов.

2. *Простота БММ*. Может создаться ложное представление, что решение все более общих задач в биологии влечет за собой усложнение математических моделей. Последнее должно диктоваться только невозможностью решения новых задач в рамках прежних моделей. Отказ от простой модели обоснован в случае, если неадекватность ее решаемой задаче доказана. Таким образом, одним из принципов моделирования является принцип простоты («бритва Оккама»).

3. *Формирование словаря БММ*. В ряде работ неудовлетворительные или противоречивые результаты являются следствием применения неадекватного словаря. Верное определение терминологии является одной из основных предпосылок успеха анализа в целом.

4. *«Сценарий» БММ*. Как правило, комплексное применение определенных моделей обусловлено не решением группы частных вопросов, а необходимостью исследования иерархии подзадач. Определение иерархии подзадач и составляет «сценарий» БММ. Естественно, что такого рода сценарий не является жесткой схемой, а создается с целью варьирования модели в зависимости от решаемой задачи.

Теперь на примерах обсудим сформулированные выше положения.

Пример 1. Поиск биологически активных соединений. В конце XIX века в рамках одномерной регрессионной модели [14] была установлена зависимость местного анестезирующего эффекта от липофильности в узком гомологическом ряду. Множество подобных соотношений, подтверждающих идею взаимозависимости биологических свойств молекул с их физико-химическими свойствами, вселяли уверенность в возможность нахождения универсальной зависимости «структура—активность» методами одномерного статистического анализа. Однако «равний мираж» вскоре рассеялся [19].

Рост исследований по количественному анализу зависимости «структура—активность» (КАСА) в 60—70 гг. был связан с применением класса моделей множественной регрессии, в которых биологическая активность определяется уже в зависимости от ряда свойств структуры. Анализ зависимости от структуры не только основной активности, но и побочного действия препарата привел к усложнению математических моделей. Еще большее усложнение их необходимо, когда ограничения на количество анализируемых активностей отсутствуют (например, в рамках анализа канонической корреляции). Далее при поиске лекарственных препаратов с использованием методов распознавания образов математические модели еще более усложняются.

Пример 1 подтверждает тезис о многомерности БММ на современном этапе.

Пример 2. Иммунный ответ организма на ревакцинации против дифтерии и столбняка. Единственная модель долгосрочного иммунитета [13] имела сложное нелинейное математическое описание. В работах [3, 4] предложена и обоснована уже линейная (!) модель иммунного ответа в фиксированные моменты времени, получившая завершённый вид [3].

Пример 3а. (Подходы к параметризации структур для КАСА). Переход к математическим моделям множественной регрессии для КАСА в 60–70 гг. обусловил необходимость адекватной кодировки — параметризации определенных количественных характеристик химических структур

Из трех возможных подходов к параметризации (эмпирического [11], полуэмпирического [15] и основанного на квантово-химических расчетах теоретического [16]) наиболее перспективным, на наш взгляд, является полуэмпирический, с которым и связан прогресс в современных исследованиях в данной области

Трудности, возникающие при практическом применении эмпирического подхода, связаны с невозможностью прогнозирования биологической активности новых членов анализируемого гомологического ряда [5]. К числу недостатков теоретического подхода относятся ограниченность и неоднозначность квантово-химического описания структуры в рамках существующих на сегодняшний день моделей [2]. Кроме того, как правило, расчеты производятся для молекул в вакууме и, тем более, не учитываются взаимодействия «препарат—рецептор», что вносит свою долю неопределенности.

Пример 3б. Эффект действия простагландинов на мозговое кровообращение. Существует множество работ, в которых утверждается сосудосуживающее действие простагландина E_1 ($PG E_1$) (см., например, [22]), в не меньшем числе работ доказывалось обратное [18]). Наконец, имеются работы, вообще отрицающие вазоактивность простагландина E_1 [10].

С целью анализа реакции сосудов различных уровней пинальной системы вводится субъективная классификация по их калибру. Авторы [17] именно этим и прабуют объяснить разнонаправленность реакции сосудов. Однако, на наш взгляд, проблема обусловлена не столько различиями в калибрах сосудов, сколько различиями в их исходных функциональных состояниях. Действительно, один и тот же сосуд при введении препарата сначала может сузиться, а потом расшириться, что при измерении его в различные моменты фиксации не исключает противоположных выводов.

Примеры 3а и 3б обосновывают необходимость формирования словаря БММ.

Остановимся на общих положениях конструирования сценария БММ.

Сложные проблемы, возникающие при исследованиях в различных областях биологии (поиск лекарственных препаратов, действие физиологически активных вещества на отдельные функциональные системы

организма и т. д.), не всегда целесообразно решать в рамках отдельно взятой, даже весьма сложной математической модели. Поэтому более перспективно искать их решение путем реализации моделей в виде технологически и логически взаимосвязанных цепочек—сценариев БММ.

Наметим примерный сценарий, служащий ядром конкретных приложений, примеры которых приводятся ниже в качестве иллюстраций.

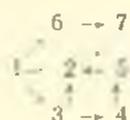


Рис. 1. Объяснение и текст.

На первом этапе (рис. 1) в силу принципа простоты апробируются линейные модели, т. е. восстанавливается вид функциональной зависимости (например, «доза—эффект»). Одним из приемов восстановления неизвестной функциональной зависимости является, например, полиномиальное сглаживание (пункт 1). Наличие адекватности линейного приближения явно упрощает последующий анализ (пункт 2—построение линейной модели). В большинстве задач биологии линейное приближение вполне приемлемо. Отсутствие адекватности можно объяснить следующим образом: а) процесс нелинейен; б) имеет место гетерогенность материала—случай, весьма часто встречающийся в биологических исследованиях.

В варианте а необходимо апробировать нелинейные модели (переход 3—2, построение нелинейной модели), естественно, более сложные. Однако зачастую нет необходимости в усложнении модели. Можно выявить более однородные подгруппы данных, внутри которых вполне возможно описание результатов в рамках линейных моделей.

В варианте б наличие гетерогенности (пункт 3) можно проверить разными способами, например, спроектировав данные на плоскость первых двух варимакс-факторов с последующим визуальным анализом проекции. Можно вычислить D^2 -расстояние Махаланобиса каждого объекта до общего среднего, что характеризует типичность объекта. Наконец, одним из общепринятых методов определения объективной классификации объектов является кластер-анализ.

При выявлении гетерогенности материала встает вопрос классификации, т. е. выделения более однородных групп в исходном материале (пункт 4), например, в рамках кластер-анализа. После выделения однородных групп следует определение решающих правил для отнесения того или иного объекта к одной из выделенных групп (пункт 5), что разрешается в рамках модели дискриминантного анализа. Далее для каждой из установленных групп возможен анализ на основе единой (например, регрессионной) модели (переход 5—2).

Отдельно можно выделить ветвь сценария, исследующую внутреннюю структуру параметров (пункты 6 и 7). Естественно, не все используемые в анализе параметры независимы. Корреляция параметров есть следствие взаимосвязанности их свойств. Еще более общий случай наблюдается, когда принципиальные свойства непосредственно не изме-

римы в эксперименте и о них можно составить представление лишь по имеющимся данным. Как правило, изучение внутренней структуры параметров, выявление непосредственно не измеримых в опыте принципиальных свойств и т. д. производится с помощью факторного анализа (пункт 6). Если же интересующее свойство описывается набором показателей, то для нахождения их связи с экспериментальными переменными вместо ряда независимых регрессионных моделей предпочтительнее использовать модель канонической корреляции (рис. 1).

Приведем примеры применения предлагаемого сценария при решении конкретных биологических проблем.

Пример 4а. Поиск синтетических лекарственных препаратов (рис. 2).

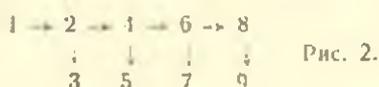


Рис. 2.

1—параметризация КАСА (параметры линейности соотношений свободных энергий); 2—доказательство адекватности КАСА для описания всего спектра (анализ канонической корреляции); 3—исследование внутренней структуры параметров (факторный анализ); 4—классификация по фармакологическому действию (кластер-анализ); 5—эталонны фармакологического действия (расчет D^2 -расстояния Махаланибиса); 6—классификация по физико-химическим свойствам (кластер-анализ); 7—эталонны по физико-химическим свойствам (D^2 -расстояние Махаланибиса); 8—взаимосвязь «класс фармакологического действия—физико-химические свойства» (линейный дискриминантный анализ); 9—дискриминантный анализ с более высокими членами.

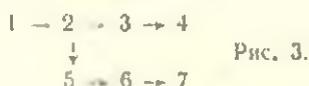
Проблема выдвигает специфические вопросы, связанные с многомерным описанием фармакологического действия и не свойственные классическому КАСА.

Во-первых, новую окраску принимает задача выброса параметризации структуры, ибо заранее не очевидно, что параметризация, удовлетворительно описывающая каждое из фармакологических свойства порознь, окажется приемлемой и при комплексном описании действия. С этим связана задача взаимосвязи физико-химических и фармакологических свойств.

Во-вторых, наиболее существенное отличие от КАСА—проблема определения понятия «предпочтительность». При анализе одного вида активности такой задачи не возникает: наиболее активное соединение, очевидно, является и наиболее предпочтительным. При одновременном анализе активности и токсичности задача усложняется—вводится индекс селективности, что сводит и этот случай к одномерному [21]. При анализе доступного спектра фармакологического действия эта задача требует специального внимания. Спецификой БММ в данном случае можно считать: 1) выбор адекватного базиса для описания структуры; 2) решение специфических задач, связанных с многопараметровым описанием действия, в частности, осмысление понятия «предпочтительность» в рамках данного подхода.

Реализация предлагаемого сценария для фармакологического действия (пять видов биологической активности) полусинтетических производных пенициллина приведена на рис. 2. Сценарий состоит из трех этапов: А. выбора адекватной параметризации и исследования взаимосвязей используемых параметров (пункты 1—3); В. определения аналога понятия «предпочтительность» (пункты 4—7); С. установления взаимосвязи фармакологического действия с физико-химическими свойствами структуры (пункты 8, 9). На этапе А доказываемся адекватность линейности соотношений свободных энергий для описания спектра действия; устанавливаются внутренние зависимости параметров [7]. На этапе В вводится многомерный аналог понятия «предпочтительность» — «препарат со схожим фармакологическим действием». Выделены основные классы фармакологического действия, определены эталоны для каждого из классов. Аналогичные результаты имеют место и в пространстве физико-химических параметров. Этап С приводит к статистически значимым соотношениям «спектр фармакологического действия — физико-химические свойства структуры», что, видимо, и является решением проблемы [8].

Пример 4б. Иммуный ответ организма на ревакцинацию против дифтерии и столбняка. Недостатки варианта БММ из [13] уже отмечались. Для решения вопросов моделирования иммунного ответа общий сценарий нами модифицирован (рис. 3).



1 — выбор анализируемых факторов (однофакторный дисперсионный анализ); 2 — исследование выбранного базиса (мультиколлинированности); 3 — внутренняя структура параметров (кластер-анализ); 4 — взаимосвязь титров по дифтерии и столбняку (анализ канонической корреляции); 5 — зависимость послепрививочного скачка (полиномиальное сглаживание);

$$\lg \frac{\text{Post}}{\text{Pre}} = a_0 + a_1 \lg \text{Pre} + a_2 (\lg \text{Pre})^2 + \dots;$$

6 — исследование зависимости уровня титра в фиксированные моменты времени (линейная регрессия $\lg \frac{\text{Titer}}{\text{Pre}} = a_0 + a_1 \lg \text{Pre}$);

7 — построение модели (множественная линейная регрессия).

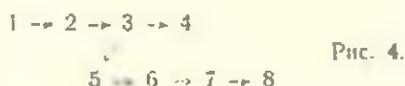
В отличие от модели Готтлиба, на этапе 1 нашего сценария анализируется влияние такого рода факторов, как пол, группа крови, сезон ревакцинации и т. д. Ортогональность расширенного базиса обсуждается на этапе 2. Исследование внутренней структуры используемых параметров проводится на этапе 3. Выявлена значительная корреляция между дифтерийными и столбнячными титрами во все моменты фиксации. Поэтому на этапе 4 решается вопрос о взаимосвязи и динамике их изменения.

Наиболее существенным моментом в построении модели является исследование скачка послевакционного титра по сравнению с довакци-
онным: $\lg \frac{\text{Post}}{\text{Pre}}$ от $\lg \text{Pre}$, который в модели Готтлиба постулируется в
виде $10^{a_0 + a_1 + \dots + a_k \text{Pre}^{a_k} (\lg \text{Pre})^{b_k}}$, причем, как уже отмечалось, столь
сложный вид зависимости не обоснован. На этапе 5 вопрос решается в
пользу линейного приближения полиномиального сглаживания.

Следующий момент моделирования долгосрочного иммунитета—
определение временной зависимости уровня антител. В предлагаемом
сценарии [4] вместо сложной эмпирической зависимости, необоснован-
но вводимой Готтлибом и соавт., рассчитываются значения титров для
определенных моментов фиксации (1,5, 6, 12, 24 и 36 месяцев). На ос-
новании результата предыдущего этапа на этапе 6 показана статисти-
ческая значимость линейности $\lg \frac{\text{Titer}}{\text{Pre}}$ от $\lg \text{Pre}$ с высокими коэффи-
циентами корреляции.

На последнем этапе на основании анализа множественной регрес-
сии строится модель иммунного ответа с отбором наиболее информатив-
тивных параметров.

*Пример 4в. Влияние вазоактивных простагландинов типа E₁ на пи-
альную систему мозгового кровообращения.* Предлагаемая для иссле-
дования влияния ПгE₁ на мозговое кровообращение модификация спе-
нария приведена на рис. 4.



1—исследование зависимости «доза—эффект» (полиномиальное сглажи-
вание); 2—построение линейной модели для описания изменения арте-
риального давления и показателей кислотно-щелочного состояния кро-
ви (множественная регрессия); 3—исследование адекватности пара-
метризации (каноническая корреляция); 4—зависимость «чистого»
действия с эффектом на фоне базального уровня простагландинов (ка-
ноническая корреляция); 5—исследование внутренней структуры пара-
метров и типичности поведения различных сосудов (факторный анализ;
D²—расстояние Махаланобиса); 6—проверка различной параметри-
зации для формирования термина «образ функционального поведения»
(линейная разделимость, D²—статистика, мощность ЛДФ); 7—объек-
тивная классификация типов функционального поведения, определение
эталонов (кластер-анализ, D²—расстояние Махаланобиса); 8—распо-
знавание функционального поведения пилальных сосудов (дискриминан-
тный анализ).

На основании данных о «колоколообразном» виде зависимости «до-
за—эффект» на этапе 1 исследуется зависимость изменения диаметров
сосудов и артериального давления (АД) в фиксированные моменты
времени. На этапе 2 строятся линейные модели для описания измене-
ния АД и параметров кислотно-щелочного состояния крови.

Линейная модель описания изменения диаметров сосудов неадекватна экспериментальным данным. Поэтому на этапе 3 исследуется вопрос о принципиальной возможности описания изменения диаметров в рамках используемой параметризации. На этапе 4 изучается взаимосвязь действия Pg E_1 на фоне базального уровня простагландинов и при ингибировании их биосинтеза индометацином. Пункт 5 выявляет внутреннюю структуру параметров, а также сосуды с типичным и нетипичным поведением. Интересно отметить, что нетипичное поведение демонстрируют как крупные, так и мелкие сосуды [6, 12].

В пункте 6 апробируются три вида параметризации «образа функционального поведения», являющегося аналогом понятия вазоактивности: а) «чистого» действия, б) действия на фоне базального уровня простагландинов, в) комплексного описания. Адекватность параметризации проверяется с помощью тестирования объективной классификации по D^2 -статистике Махаланобиса или по интуитивно ожидаемому критерию — линейной разделимости классов. В пункте 7 дается объективная классификация функционального поведения пинальных сосудов и выявляются эталоны для всех пяти групп поведения. Наконец, на этапе 8 с помощью пошагового дискриминантного анализа находят решающие правила для описания (и прогноза) поведения индивидуальных сосудов в течение 20 мин после введения определенной дозы Pg E_1 . Основные результаты применения сценария приведены в [6, 12].

Пример 4г. Исследование влияния протекторов на эффект действия мутагенов. На протяжении пятнадцати лет отлаживается методологический поиск математических подходов к моделированию механизмов действия протекторов на основе изучения кривых «концентрация мутагена — цитогенетический эффект» в отсутствие протектора и при его введении в культуру клеток.

Эмпирически отработанный сценарий полностью укладывается в предлагаемую нами схему. На этапе 1 выясняется возможность описания концентрационных кривых в рамках линейного регрессионного анализа. Этап обусловлен тем, что, как правило, зависимость «концентрация мутагена — эффект» для доли aberrантных клеток и количества разрывов в клетках имеет нелинейную форму. Поэтому необходимо путем определенного масштабирования линеаризовать кривую (например, один из возможных путей указан в [9]).

На этапе 2 тестируются одинаковые математические модели для вариантов, обработанных мутагеном с добавлением протектора и без него. Таким образом, на данном этапе решается вопрос о наличии влияния протектора на механизм действия мутагена. На этапе 3 исследуется внутренняя структура параметров, описывающих влияние протектора на эффект мутагена. Этап реализован в рамках модели кластер-анализа. На этапе 4 построение количественной модели влияния протектора на эффект действия мутагенов осуществляется в рамках множественного регрессионного анализа с пошаговым выбором информативных параметров [1].

Приведенные примеры, естественно, не исчерпывают имеющегося даже у авторов опыта применения предлагаемого подхода и подтверждают наличие достаточно широкого спектра применений универсального сценария БММ. В дальнейшем, очевидно, еще больше расширится спектр приложений, что будет способствовать проверке, уточнению и обогащению сценария БММ, наполняя его новым содержанием.

ЛИТЕРАТУРА

1. Арццунян Р. М. Модификация хлмического мутагенеза в клетках человека. Ереван, 1985.
2. Ландиу М. А. Хим.-фарм. ж., 9, 112—114, 1982.
3. Манаслян Е. В. Автореф. канд. дисс., М., 1984.
4. Ордуханян А. А., Манаслян С. В. Биолог. ж. Армении, 36, 7, 592—596, 1983.
5. Ордуханян А. А. Автореф. канд. дисс., М., 1979.
6. Ордуханян А. А., Налбандян С. Г. В кн.: I Всесоюзн. конф. «Фитнохимия, патология и фармакология мозгового кровообращения». Ереван, 131—132, 1984.
7. Ордуханян А. А., Ландиу М. А., Мнджоян Ш. Л., Тер-Захарян Ю. З. Хим.-фарм. ж., 2, 65—70, 1980.
8. Ордуханян А. А., Саркисян А. С., Ландиу М. А., Мнджоян Ш. Л., Тер-Захарян Ю. З. Хим.-фарм. ж., 1, 63—66, 1983.
9. Чеботарев А. Н., Яковенко К. Н. Генетика, 10, 8, 150—157, 1974.
10. Carlson I. A., Ekelund T. G., Oro L. L. Acta Med. Scand., 183, 423—430, 1968.
11. Free S. N., Willson J. M. J. Med. Chem., 7, 395—399, 1964.
12. Gabrielyan E. S., Amroyan E. A., Nalbandian S. G., Ordukhanyan A. A. In: Application of new electronic instruments in chemical and experimental pharmacology, Ithringen, 102—105, 1984.
13. Gottlieb S., Martin M., MacLaughlin F. Y., Panam R. Z., Levin L., Edsall G. Amer. J. Epidemiol., 85, 2, 207—219, 1967.
14. Drum-Braun A., Frazer T. Trans. Roy. Soc. Ed'nb., 25, 151, 1869.
15. Hansch C., Leo A., Elkins D. J. Chem. Soc., 14, 2, 57—69, 1974.
16. Kier L. B. In: Fundamental concepts in drug-receptor interactions, N. Y., 1970.
17. Mackenzie E. T., Stranguard S., Graham D. J. Circulation Res., 39, 1, 33—41, 1976.
18. Nakano J., Chang A. C., Fisher R. G. J. Neurosurgery., 38, 32—39, 1973.
19. Parascandola J. Pharmacy in History, 18, 1, 3—10, 1971.
20. Pell J. Biom. J., 21, 1, 57—70, 1979.
21. Tichy M. In: Quantitative structure-activity analysis, 359, Berlin, 1978.
22. Yamamoto Y. I., Wolfe L. S., Hodges C. P., Fetndel W. Excerpta Medica Intern. Congress Series, 293, 56, 1973.

Поступило 15.IV 1986 г.