

## СОВРЕМЕННАЯ МОДЕЛЬ ОБЪЕКТИВНОЙ ЗАКОНОМЕРНОСТИ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ПРИ ПРОВЕРКЕ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Основой теории проверки статистических гипотез, которая широко используется при анализе и проектировании информационно-измерительных комплексов различного назначения (обнаружение сигналов, измерение их параметров, передача, хранение и обработка информации), является методика проверки простой гипотезы против простой альтернативы.

В лемме Неймана-Пирсона зафиксированы (по состоянию на 1928 – 1933 гг.) научные представления об объективной закономерности извлечения информации при проверке простой гипотезы против простой альтернативы. Именно эти представления и их следствия отражены в литературе и воплощены в самых современных системах извлечения, передачи, хранения и обработки информации, в частности, при организации ее защиты.

При использовании современной модели объективной закономерности извлечения информации при проверке статистических гипотез может быть достигнуто повышение эффективности средств защиты информации за счет более качественного распознавания состояний угрозы атаки или проявлений ее начальных последствий, а также уменьшение числа случаев ложных тревог. Последние могут наносить прямой и существенный эксплуатационный или экономический ущерб, так как непосредственно связаны с необходимостью изоляции, существенно более тщательного тестирования или даже преждевременного обновления как программных так и аппаратных средств поддержки информационных процессов.

В брошюре [1] предложена более общая модель объективной закономерности извлечения информации при проверке простой гипотезы против простой альтернативы, отражающая наши знания по состоянию на 1997 – 1999 гг. Форма связи между моделями 1928 – 1933 гг. и 1997 – 1999 гг. отвечает принципу соответствия, который был сформулирован Н.Бором в 1913 г. Показано, что лемма Неймана-Пирсона корректно отображает действительность лишь при отношениях сигнал/шум ( $c/\omega$ ) равном 0,5 и менее.

Использование алгоритмов, опирающихся не на модель 1928 – 1933 гг., а на модель 1997 – 1999 гг., позволяет (по крайней мере теоретически) получить те же вероятности ошибок 1-го и 2-го рода при меньших объемах выборок. Так в [2] на одном из примеров показано, что при отношении  $c/\omega=1,0$  объем выборки может быть уменьшен почти на 10%, при  $c/\omega=1,5$  – на 30%, при  $c/\omega=2,0$  – в 3 раза, а при  $c/\omega=2,5$  – в 10 раз.

Как известно, “законы науки являются отражением законов природы. Они открываются и формулируются учеными и, следовательно, представляют собой наши знания о законах природы. ..., иными словами, они являются идеальными (мысленными, понятийными) моделями законов природы” [3].

Раскрыть содержание того или иного объективного закона и сформулировать соответствующий ему научный закон вовсе не просто. История развития науки показывает, что открытие закона природы “обычно происходит не сразу, не до конца, а в форме неполного, приближенного, относительного знания. Лишь в дальнейшем, на каждой последующей ступени развития науки, смысл и содержание объективного закона раскрывается все глубже и полнее, а формулировка соответствующего научного закона постепенно уточняется и становится все более адекватной отражаемому им объективному закону. Это неполное соответствие между научным и объективным законами обусловлено прежде всего сложной структурой самой действительности” [3]. Так, например, сначала был сформулирован закон сложения скоростей движущихся тел в рамках классической механики и лишь много лет спустя в релятивистской механике этот закон был существенно дополнен,

хотя и классическая механика и релятивистская механика отображают наши знания об одном и том же объекте.

Следует отметить, что информация является относительно новым объектом, свойства которого изучаются естествознанием, и нет ничего необычного в том, что не все ее свойства изучены достаточно полно.

Лемма Неймана - Пирсона, или как еще ее называют, фундаментальная лемма математической статистики [4] является одним из примеров научного закона, сформулированного на языке математической статистики, которому соответствует объективная закономерность извлечения информации из окружающей среды (закономерность передачи информации при наличии помех). "... здесь справедливо говорить не о решении математической задачи, а о математическом решении задачи, объектом которой является физическая (материальная) система" [5]. О степени фундаментальности этой закономерности можно судить по разнообразию предметных областей в которых она проявляется. В литературе этот научный закон рассматривается скорее только как лемма математической статистики, а не как модель, отображающая объективную закономерность передачи информации, которой подчиняется широкий класс объектов. Но именно эту объективно существующую закономерность в области передачи информации открыли и описали Нейман и Пирсон в 1928 году, а в 1933 году привели и математическое доказательство своей правоты [6].

Фундаментальная лемма математической статистики (как модель объективной закономерности извлечения информации из окружающей среды) является теоретическим обоснованием оптимальности многих алгоритмов обнаружения сигналов и оценки их параметров. При этом имеются в виду не только алгоритмы, которые реализованы в технических системах, но и алгоритмы, изложенные в инструкциях и других руководящих документах, определяющих порядок и логику действий тех или иных должностных лиц при решении практических задач, которые в информационном плане можно рассматривать как задачи обнаружения сигналов и оценки их параметров. Алгоритмы обнаружения сигналов и оценки их параметров, в основу которых положен критерий отношения правдоподобия (байесовский, максимума апостериорной вероятности, Неймана - Пирсона и др.), считаются оптимальными. Такая точка зрения и ее следствия в настоящее время являются общепринятыми, и они нашли свое отображение в учебниках, монографиях, справочниках, энциклопедиях и отображены в литературе в виде описаний оптимальных алгоритмов обнаружения, измерения параметров и алгоритмов обработки разнообразной информации, которые воплощены в самых современных технических системах передачи, воспроизведения (хранения) и обработки информации. Можно предположить, что обработка информации в биологических системах, которые формировались под действием объективных факторов, а не под влиянием тех или иных научных теорий, в большей степени соответствует объективно существующей закономерности извлечения информации из окружающей среды, чем обработка информации в технических и социальных системах.

Формулировка научного закона, приведенная в [7], соответствует той же самой объективной закономерности, о которой идет речь в лемме Неймана - Пирсона. Но модель этой объективной закономерности в [7 и 8] существенно отличается от модели той же закономерности, приведенной в лемме Неймана - Пирсона.

Что же общего в этих двух моделях, время появления которых разделяет интервал почти в 70 лет, и чем они отличаются друг от друга?

Обе эти модели могут быть применены для рассмотрения элементарного акта передачи информации при наличии помех - процесса принятия решения о том, какой именно символ из двух возможных ("0" или "1") был передан, на основании наблюдения текущих значений некоторой физической величины (переносчика сигнала) на протяжении некоторого временного интервала.

Принятие решения о каждом переданном символе производится по результатам нескольких независимых наблюдений (серии измерений), и наблюдателю известны моменты начала и окончания передачи каждого символа.

Результат каждого наблюдения зависит не только от переданного символа, но и от случайного значения помехи.

Наблюдателю известны условные плотности распределения значений наблюдаемой физической величины при передаче символа "0" и при передаче символа "1".

Процедура принятия решения в обеих моделях состоит из трех этапов:

этап 1 - каждому значению  $x$  наблюдаемой физической величины, зафиксированному при наблюдении, по определенному правилу ставится в соответствие значение  $y$  другой величины;

этап 2 - по результатам серии наблюдений одного временного интервала, на протяжении которого передается только один из двух возможных символов, вычисляется среднее значение (или сумма значений)  $y$ ;

этап 3 - полученное среднее значение  $y_{cp}$  (или сумма значений) сравнивается с некоторым пороговым значением, которое устанавливается (принимается) до начала серии наблюдений, исходя из допустимых вероятностей ошибок первого и второго рода. Результат сравнения однозначно определяет решение ("передан символ 0" или "передан символ 1"), которое должно быть принято по серии наблюдений.

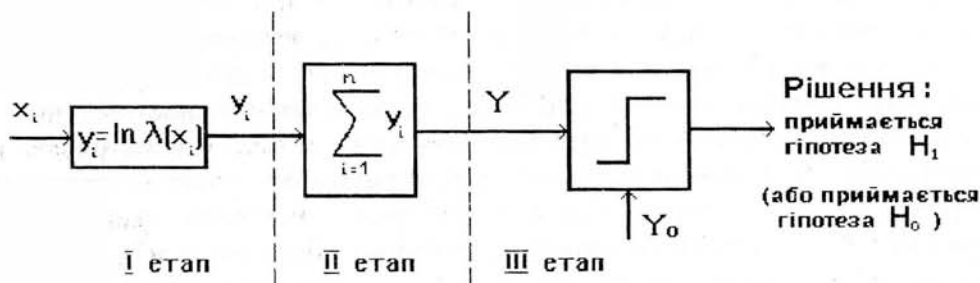


Рис. 1. Структурно-логическая схема правила принятия решения при проверке простых статистических гипотез  $H_0$  и  $H_1$

Иначе говоря, во время первого этапа процедуры принятия решения, наблюдатель переходит от исходной наблюдаемой физической величины ( $x$ ), к другой, вспомогательной (промежуточной, вторичной) наблюдаемой величине ( $y$ ), по совокупности значений которой в заданном временном интервале он и принимает решение (второй и третий этапы) о том, какой именно символ из двух возможных символов был передан в заданном временном интервале.

Сравниваемые модели различаются между собою лишь правилом, по которому каждому значению  $x$  исходной наблюдаемой физической величины ставится в соответствие значение  $y$  вспомогательной величины во время первого этапа процедуры принятия решения. То есть, сравниваемые модели различаются только правилом перехода от исходной наблюдаемой физической величины ( $x$ ), к другой, вспомогательной наблюдаемой величине ( $y$ ).

В модели Неймана - Пирсона правило такого перехода описывается формулой (1) [9]:

$$y = \ln[W_1(x)/W_0(x)], \quad (1)$$

где:  $y$  - вспомогательная наблюдаемая величина;

$W_1(x)/W_0(x)$  - отношение правдоподобия;

$W_1(x)$  - условная плотность распределения значений наблюдаемой физической величины  $x$  при передаче символа "1";

$W_0(x)$  - условная плотность распределения значений наблюдаемой физической величины  $x$  при передаче символа "0".

В модели 1997 года, описанной в [7, 8], сформулированы требования, которым должна удовлетворять вспомогательная наблюдаемая величина ( $y$ ) для достижения максимальной эффективности передачи информации при наличии помех. Эти требования состоят в том, чтобы условные распределения вспомогательной наблюдаемой величины ( $y$ ) соответствовали минимуму выражения (2) [8].

$$\{(\sigma_{0y}z_F + \sigma_{1y}z_D)/(M_1 - M_0)\} \rightarrow \min, \quad (2)$$

где:  $M_1 > M_0$ ;

$\sigma_{0y}$  - среднеквадратичное отклонение значений  $y$  при приеме символа "0";

$\sigma_{1y}$  - среднеквадратичное отклонение значений  $y$  при приеме символа "1";

$M_0$  - среднее значение (математическое ожидание)  $y$  при приеме символа "0";

$M_1$  - среднее значение  $y$  при приеме символа "1";

$z_F$  - коэффициент, значение которого зависит от допустимых значений вероятностей ошибок 1-го рода и функции распределения накопленных значений  $y$  при приеме символа "0" [10];

$z_D$  - коэффициент, значение которого зависит от допустимых значений вероятностей ошибок 2-го рода и функции распределения накопленных значений  $y$  при приеме символа "1" [10];

Как видно из (1), в модели Немана - Пирсона значения вспомогательной наблюдаемой величины ( $y$ ) зависят только от отношения  $W_1(x)/W_0(x)$  в точке  $x$ . В модели 1997 года (2) значения вспомогательной наблюдаемой величины ( $y$ ) зависят не только от текущего значения исходной наблюдаемой физической величины ( $x$ ), но и от шести параметров, два из которых ( $z_F$ ,  $z_D$ ) определяются допустимыми вероятностями ошибок 1-го и 2-го рода.

Зависимость между значениями  $z_F$  и  $z_D$ , с одной стороны, и значениями вероятностей ошибок 1-го и 2-го рода - с другой, можно показать с помощью следующих соотношений:

$$\alpha = 1 - \Phi_0(z_F), \quad \beta = \Phi_1(z_D),$$

где:  $\alpha$  - допустимая вероятность ошибок 1-го рода;

$\beta$  - допустимая вероятность ошибок 2-го рода;

$\Phi_0(z_F)$  - нормированная функция распределения накопленных

значений у на выходе накопителя при приеме символов "0";  
 $\Phi_1(z_D)$  - нормированная функция распределения накопленных  
 значений у на выходе накопителя при приеме символов "1".

Обычно функции  $\Phi_0$  и  $\Phi_1$  с достаточной для практики точностью описываются функцией закона нормального распределения.

$$z_F = V_0 / \sigma_0, \quad z_D = V_1 / \sigma_1.$$

где:  $V_0$  - превышение порогового уровня над математическим  
 ожиданием накопленных значений у на выходе накопителя при приеме  
 символов "0" ;

$V_1$  - превышение над пороговым уровнем математического ожидания  
 накопленных значений у на выходе накопителя при приеме символов "1"  
 ;

$\sigma_0$  - среднеквадратичное отклонение накопленных  
 значений у при приеме символов "0" ;

$\sigma_1$  - среднеквадратичное отклонение накопленных  
 значений у при приеме символов "1".

В 1913 году Н. Бор сформулировал важный методологический принцип, получивший название принципа соответствия. "Согласно этому принципу всякая более общая теория включает в себя старую теорию; старая теория получается из новой при предельном переходе к определенным значениям определяющих ее параметров. Так, законы квантовой механики переходят в законы классической физики при условии, что можно пренебречь значением кванта действия, а законы теории относительности переходят в законы классической механики при условии, что скорости движущихся тел или частиц малы по сравнению со скоростью света. Следовательно, новая теория не отменяет старую, а лишь уточняет границы, в которых эта старая теория продолжает действовать" [1].

Как показали результаты исследований [11], при значении параметра с/ш близком к 0 (0.5 и менее), обобщенный критерий [12], который учитывает закономерность изменения эффективности накопления сигнала двоичного кода, совпадает по своим результатам с критерием отношения правдоподобия, хотя значения этих критериев вычисляются по совершенно разным правилам. Зато при значении параметра с/ш порядка 1 и более эти разные правила приводят и к разным результатам: с ростом отношения сигнал/шум эффективность алгоритмов, которые учитывают закономерность изменения эффективности накопления сигнала двоичного кода, существенным образом возрастает, по сравнению с эффективностью существующих оптимальных алгоритмов, основанных на использовании критерия отношения правдоподобия.

Таким образом, новая модель объективного закона, которая предложена в 1997 году [8], позволила уточнить границы [11], в которых модель объективного закона, предложенная Нейманом и Пирсоном в 1928 году [4], сохраняет свою корректность. Эта граница позволяет определить те предметные области, в интересах которых (в первую очередь) нужно развернуть научные направления по совершенствованию алгоритмов обработки информации - приведению их в соответствие с моделью 1997 года.

"... развитие науки идет по линии установления все более общих и, следовательно, информативно более емких законов, включающих в себя менее общие законы" [3]. Именно таким более общим законом является "закономерность изменения эффективности накопления сигнала двоичного кода" [7].

Предвидение (предсказание) результата того или иного реального действия является необходимой частью любой целенаправленной деятельности, которая включает операцию

преднамеренного выбора. И этот выбор в любом случае может опираться только на прошлый опыт независимо от того, в каком виде этот прошлый опыт выступает - выступает ли он в виде обычного житейского "здорового смысла" или в виде самой современной научной теории, обобщающей наши знания в той или иной предметной области (метеорология, земледелие, космонавтика, экономика, политика).

Применение нового научного представления об основной закономерности извлечения информации из окружающей среды в процессе решения самых разнообразных прикладных задач позволит свести к минимуму потери от влияния помех на процесс принятия решения. Но для этого необходимо внести соответствующие коррективы в методики решения более сложных задач теории проверки статистических гипотез, чем задача проверки простой гипотезы против простой альтернативы.

Исходя из изложенного, можно с полным правом утверждать, что к современным информационным технологиям можно отнести только те, в которых на современной технической базе реализованы современные научные знания.

### ЛИТЕРАТУРА

1. Ручкін В.О. Закономірність добування інформації. К.: Товариство "Знання" України, 2002. - 45 с.
2. Ручкін В.О. Методика порівняльної оцінки ефективності алгоритмів виявлення пачки некогерентних імпульсів на фоні флюктуаційних шумів // Досвід навчально-методичної роботи Київського військового інституту управління та зв'язку. Науково-методичний збірник. - Вип. 5. - К.: КВІУЗ, 1996. - С. 203 - 206.
3. Друянов Л.А. Законы природы и их познание. - М.: Просвещение, 1982. - 112 с.
4. Неймана-Пирсона лемма // Математическая энциклопедия. Т. 3. (Коо - Од) / Гл. ред. И.М. Виноградов. - М.: Советская Энциклопедия, 1982. - 1184 стб.
5. Госсен И.И., Колотушина С.П., Тыминский В.Г. Физику о научном открытии. - Томск: Томский университет, 1984. - 103 с.
6. Отношения правдоподобия критерий // Математическая энциклопедия. Т. 4. (Ок - Сло) / Гл. ред. И.М. Виноградов. - М.: Советская Энциклопедия, 1984. - 1216 стб.
7. Ручкин В.А. Закономерность изменения эффективности накопления сигнала двоичного кода. - К.: Киев. воен. ин-т. упр. и связи. Деп. в ГНТБ Украины 01.09.99, №235 - УК99.
8. Ручкин В.А. Методика автоматизированного нахождения оптимального решения задачи проверки простой гипотезы против простой альтернативы. - К.: Киев. воен. ин-т. упр. и связи. Деп. в ГНТБ Украины 06.02.97, №154 - УК97.
9. Ван Трис Г. Теория обнаружения, оценок и модуляции. Т. 1. - М.: Сов. радио, 1972. - 744 с.
10. Ручкин В.А. Номограмма для определения количественных соотношений между вероятностью ложной тревоги и вероятностью правильного обнаружения сигнала // Труды КВИРТУ - К.: Киевское высш. инж. р-т. училище ПВО, 1968. - N44. - С. 57-61.
11. Ручкин В.А. Скорректированное отношение правдоподобия и эффективность его использования при проверке простой гипотезы против простой альтернативы. - К.: Киев. воен. ин-т. упр. и связи. Деп. в ГНТБ Украины 12.06.97, №359 - УК97.
12. Ручкин В.А. Статистический критерий оптимальной обработки сигналов для нового поколения систем обнаружения и измерения. - К.: Киев. воен. ин-т. упр. и связи. Деп. в ГНТБ Украины 14.12.98, №488 - УК98.

Поступила 18.07.2002