ТЕОРИЯ ПЕРЕДАЧИ СООБЩЕНИЙ

(Труды Третьей международной конференции)

СБОРНИК СТАТЕЙ

Под редакцией
чл.-корр. АН СССР
В. И. СИФОРОВА

ИЗДАТЕЛЬСТВО
ИНОСТРАННОЙ ЛИТЕРАТУРЫ
МОСКВА, 1957

ПРЕДИСЛОВИЕ

Теория передачи сообщений (теория информации) изучает общие закономерности, присущие как самим сообщениям, так и их передаче при наличии помех. В силу своей весьма большой общности эта теория имеет большое значение для самых разнообразных областей науки и техники. К таким областям относятся, например, радиосвязь, проводная связь, радиолокация, радиотелеуправление, автоматика и телемеханика, техника вычислительных машин, телевидение, физиология, языкознание и т. д.

Теория передачи сообщений является одной из основ современной техники связи, и ее применение к этой области, несомненно, позволит значительно продвинуть решение таких сложных проблем, как повышение эффективности и надежности связи в условиях помех.

Начало общей теории передачи сообщений было положено в работе В. А. Котельникова „О пропускной способности „эфира“ и проволоки в электросвязи“, написанной в 1933 г., в которой автор сформулировал ряд положений и теорем, важных для последующего развития теории передачи сообщений. В 1935 г. была опубликована работа Д. В. Агеева, доказавшая принципиальную возможность увеличения числа каналов радиосвязи в заданном диапазоне частот. В 1947 г. вышла в свет новая работа В. А. Котельникова, где было введено понятие о потенциальной помехоустойчивости. Кроме того, в ней были рассмотрены еще и способы улучшения защиты линий радиосвязи от действия помех.

Наиболее бурное развитие теории передачи сообщений наблюдалось после ставших теперь уже классическими работ американского ученого Клода Шеннона (1947—1948 гг.). За последние восемь лет эта теория стала зрелой и весьма разветвленной отраслью науки, причем общее число научных работ в мировой литературе, посвященных этой теории и ее разнообразным приложениям, уже давно превысило тысячу.

Обсуждению достигнутых результатов и путей решения различных трудных проблем теории передачи сообщений был посвящен ряд международных совещаний ученых различных стран.

В сентябре 1955 г. в Лондоне состоялся третий международный симпозиум по теории передачи сообщений, на котором мне довелось присутствовать. В работе этой конференции приняли участие около 250 ученых Англии, СССР, США, Франции, Голландии, Италии, Федеративной Республики Германии, Швеции, Швейцарии, Дании и других стран. Участники этого совещания обсудили около 50 докладов по различным вопросам теории передачи сообщений и ее применения к электро- и радиосвязи, теории математических машин, физиологии, языкознанию и т. д.

В настоящий сборник включены наиболее интересные из прочитанных на этом совещании зарубежными учеными доклады по общей теории передачи сообщений и теории кодирования, а также по применению этих теорий к технике связи и некоторым проблемам структуры языка и статистики.

Ценность публикуемых докладов заключается не только в том, что в них изложены новые результаты, но также и в том, что в них сформулированы вопросы и проблемы, которые должны быть как можно скорее разрешены совместными усилиями ученых. Попутно нельзя не отметить, что ряд выдвинутых в этих докладах положений еще остается спорным, подлежит дискуссии и требует критического подхода со стороны читателя.

В. И. Сифоров.

I. ОБЩИЕ ВОПРОСЫ ТЕОРИИ ПЕРЕДАЧИ СООБЩЕНИЙ

О НЕКОТОРЫХ МЕРАХ „ИНФОРМАЦИИ", ИСПОЛЬЗУЕМЫХ В СТАТИСТИКЕ [1)]

Шютценбергер М. П.

I

Хорошо известно, что понятия информации, которыми пользуются в статистике и в теории связи, различаются как по своему формальному выражению, так и по существу.

В статистике, где задача состоит в оценке значения неизвестного параметра θ путем наблюдения состояния ξ физической системы, априорная вероятность которого $P(\xi/\theta)$ зависит от θ, приходят к выражению

$$F = \sum \left[\frac{\partial P(x/\theta)}{\partial \theta} \right]^2 \frac{1}{P(x/\theta)},$$

где суммирование производится по всем возможным значениям x величины ξ. Целое семейство теорем [2, 4, 6] связывает при различных условиях регулярности величину F с нижней границей дисперсии разности $\theta - \hat{\theta}$ между истинным значением параметра θ и его оценкой $\hat{\theta}$.

С другой стороны, в теории связи количество информации о самой величине ξ обычно оценивается выражением

$$H = -\sum P(x) \log P(x).$$

Интересно, что именно это выражение привлекло к себе значительное внимание, а не более старое выражение для F, введенное Рональдом Фишером еще в 1921 г. [3] и лишь мимоходом рассматривавшееся специалистами по связи.

Однако величины F и H не являются единственными мерами информации по отношению к тому, что содержится в эксперименте, включающем априорную вероятность. Вторая основная задача статистики — проверка на основании информации о величине ξ, какая из гипотез $\theta = \theta_0$ или $\theta = \theta_1$ верна, — естественно приводит к выражению

$$W(\theta_0, \theta_1) = \sum P(x/\theta_1) \log \frac{P(x/\theta_0)}{P(x/\theta_1)}.$$

[1)] Schützenberger M P, On Some Measures of „Information" Used in Statistics (Paris, 1955).

Вальд
А. ~~Уолд~~ [9] показал, что, каков бы ни был метод проверки (метод последовательного анализа или какой-либо другой), математическое ожидание числа независимых испытаний, необходимых для достижения заданного доверительного уровня, не может быть меньше K/W, где K зависит от вероятности ошибки, определяющей доверительный уровень. Следовательно, W можно назвать *мерой доставляемой через* ξ *информации о дилемме* $\theta = \theta_0$ *или* $\theta = \theta_1$.

Действительно, между F, H и W существует очень тесная связь.

Следуя Бартлетту [1], рассмотрим видоизмененное выражение

$$H' = -\sum P(x/\theta) \log P(x/\theta + \varepsilon)$$

и предположим, что $\log P(x/\theta + \varepsilon)$ можно разложить в ряд по возрастающим степеням ε. После некоторых упрощений получим

$H' = H + \varepsilon^2 F +$ члены высшего порядка по ε.

Аналогично, если $\theta_0 = t + \varepsilon$ и $\theta_1 = t - \varepsilon$, где ε бесконечно мало, то можно показать, что случайная переменная

$$z(x) = \log \frac{P(x/\theta_0)}{P(x/\theta_1)},$$

математическое ожидание которой при $\theta = \theta_1$ есть $W(\theta_0, \theta_1)$, имеет распределение со средним значением $2\varepsilon^2 F$ и дисперсией $4\varepsilon^2 F$, с точностью до членов высшего порядка по ε. Более общие соотношения между F и W недавно исследовал Каллбэк [5].

Цель настоящей работы, так же, как и статьи [8], — показать, что эти аналогии имеют глубокие корни в самой природе того, что мы готовы называть „мерой информации". Основной принцип аксиоматики, которую мы попытаемся дать, представляет собой в большей или меньшей степени развитие подхода Вудворда [10] к той же задаче. А именно, производя полное определение ξ, можно остановиться на промежуточном уровне и получить полную информацию путем сложения:

а) члена, соответствующего информации, полученной до этого момента;

б) члена, соответствующего последующей информации, взвешенной подходящими условными вероятностями.

Однако мы ограничимся постулированием этого „принципа Гюйгенса" только для такого промежуточного момента, который наверняка исключает некоторые возможности, и не будем требовать справедливости принципа для всех промежуточных моментов, как это имеет место у Вудворда.

В таком ослабленном виде задача допускает чисто алгебраическое рассмотрение, которое дает, кроме „обычного" H, выражения для F и W как частные случаи более полного решения; это более полное решение можно выразить в явном виде при определенных условиях регулярности.

Для простоты доказательство разбито на две части. Первая часть (условие I и теорема I) применима не только к информации. Она приводит к абстрактному эквиваленту принципа разделения переменных. Вторая часть (условие II и теорема II) определяет специфический характер информации, т. е. вводит функцию $\log P(x)$.

Условия регулярности III и IV, по-видимому, можно ослабить введением другого постулата, выполненного для H, F и W, а именно, условия неотрицательности информаций. Здесь нужны дальнейшие исследования, о чем мы только упомянем вместе с возможностью распространить теорему I на предельный случай при соответствующих ограничениях.

II

Прежде всего заметим, что мы ищем не меру информации, получаемой из *одного* данного результата наблюдений, а меру того количества информации, которое может быть получено *в среднем* с помощью данных средств наблюдения.

Чтобы представить себе общую картину, рассмотрим физическую систему, состояние которой ξ неизвестно. Для простоты будем считать, что ξ может принимать только одно из конечного множества E значений $x, y, z, \ldots$ Практически такое квантование можно предполагать всегда, если бы даже ξ было непрерывной переменной, так как любое реальное измерение может быть сделано только с конечной точностью.

Априорные вероятности, с которыми ξ может принимать состояния $x, y, \ldots$, будем обозначать через $P(x)$, $P(y), \ldots$, считая при этом, что они являются функциями некоторых неизвестных параметров, обозначаемых через θ.

По отношению к физической системе наблюдатель Ω_i характеризуется степенью точности, с которой он способен распознать ξ. Например, пусть ξ — числовая переменная с возможными значениями $E\{0, 1, 2, 3\}$. Наблюдатель Ω_1 может оказаться не в состоянии узнать о ξ больше, чем то, является ли оно нулем или нет; другой же наблюдатель Ω_2 может оказаться в состоянии лишь установить, является ли ξ четным или нет и т. д.

Поэтому каждому наблюдателю Ω_i соответствует отношение эквивалентности ρ_i между возможными состояниями ξ, или, другими словами, разбиение множества E на такие непересекающиеся подмножества (X), (Y), $(Z), \ldots$, что Ω_i не может различить два состояния, когда они принадлежат к одному и тому же компоненту (X), $(Y), \ldots$, т. е. к одному и тому же „классу эквивалентности" отношения ρ_i[1]).

[1]) Об отношениях эквивалентности см. П. С. Александров, Введение в общую теорию множеств и функций, ОГИЗ — Гостехиздат, М. — Л., 1948, гл. 1, §§ 3, 5. — *Прим. ред.*

Между отношениями эквивалентности на E существует обычное отношение частичного упорядочения $\rho' < \rho$ (ρ' „тоньше" чем ρ, т. е. каждый класс отношения ρ' содержится в каком-либо классе отношения ρ). Оно означает, что наблюдатель Ω' способен обнаружить все те различия между состояниями, что и Ω. Далее, если $\rho' < \rho$ и если X есть класс отношения ρ, то мы обозначим через $\rho'[X]$ („сужение ρ' к X") отношение эквивалентности, индуцируемое отношением ρ' в подмножестве X множества E. Если для некоторого подмножества X $\rho'[X] = \rho''[X]$, мы будем писать $\rho' \equiv \rho''(X)$ („ρ' и ρ'' тождественны на X"). Вооружившись этими понятиями, мы можем теперь сравнить наблюдателей или, вернее, пары наблюдателей.

Определение. *Две пары отношений эквивалентности* (ρ_i, ρ_j) *и* (ρ_k, ρ_l), *где* $\rho_i < \rho_j$ *и* $\rho_k < \rho_l$, *будем считать находящимися в отношении* $\sim$ *тогда и только тогда, когда существует разбиение* E *на два таких непересекающихся подмножества* E' *и* E'', *что*

$$\rho_i \equiv \rho_k(E'), \quad \rho_j \equiv \rho_l(E'), \quad \rho_i \equiv \rho_j(E''), \quad \rho_k \equiv \rho_l(E'').$$

Например, пусть $E = \{a, b, c, d, e, f, g\}$ и

$$\rho_i = (abc)(d)(e)(fg); \quad \rho_j = (abc)(d)(efg);$$
$$\rho_k = (ab)(cd)(e)(fg); \quad \rho_l = (ab)(cd)(efg).$$

Тогда $(\rho_i, \rho_j) \sim (\rho_k, \rho_l)$. Действительно, положив $E' = (efg)$ и $E'' = (abcd)$, получим:

$$\rho_i[E'] = \rho_k[E'] = (e)(fg); \qquad \rho_j[E'] = \rho_l[E'] = (efg);$$
$$\rho_i[E''] = \rho_j[E''] = (abc)(d); \quad \rho_k[E''] = \rho_l[E''] = (ab)(cd).$$

Легко показать, что символ $\sim$ является опять *отношением эквивалентности для множества всех упорядоченных пар, составленных из отношений эквивалентности.* Для этого достаточно заметить, что E'' однозначно определяется из ρ_i и $\rho_j (\rho_i < \rho_j)$ как объединение тех классов, которые являются одновременно классами отношения ρ_i и отношения ρ_j. Заметим, между прочим, что если также $\rho_j < \rho_l$, то $\rho_i < \rho_k$ и $(\rho_i, \rho_k) \sim (\rho_j, \rho_l)$. Поэтому отношение $\sim$ выглядит как абстрактная формулировка отношения равенства между двумя дробями.

Теперь для заданного конечного E рассмотрим множество R всех отношений эквивалентности на E и функцию $f(\cdot)$, отображающую R в некоторую аддитивную группу $\mathfrak{A}$.

Определение. *Функция* $f(\cdot)$ *будет называться нормировкой на* R, *если каждая четверка отношений* $(\rho_i, \rho_j) \sim (\rho_k, \rho_l)$ *влечет* $f(\rho_i) - f(\rho_j) = f(\rho_k) - f(\rho_l)$.

Теорема I. *Любая нормировка* $f(\cdot)$ *множества всех отношений эквивалентности на конечном* E *может быть записана в виде*

$f(\rho) = \sum g(X)$, *где суммирование производится по всем классам* X *отношения* ρ.

Доказательство. Для любого отношения эквивалентности с K классами $\rho = (X)(Y)\ldots(T)$ рассмотрим два других отношения: а) ρ_X — наиболее „тонкое" среди отношений эквивалентности, допускающих класс (X); б) $\rho_{\overline{X}}$ — наиболее „тонкое" среди отношений эквивалентности, допускающих классы $(Y)\ldots(T)$. Пусть ρ_0 — наиболее „тонкое" среди всех отношений эквивалентности, тогда $(\rho_0, \rho_X) \sim \sim (\rho_{\overline{X}}, \rho)$, т. е. если $f(\cdot)$ есть оценка, то $f(\rho) = f(\rho_X) + f(\rho_{\overline{X}}) - f(\rho_0)$. Следовательно, если теорема доказана для всех отношений эквивалентности не более чем с $K - 1$ классами, то она доказана для отношений с K классами, так как

$$f(\rho) = [\sum_{x \in X} g(x) + g(Y) + g(Z) + \ldots + g(T)] +$$
$$+ [g(X) + \sum_{x \in E - X} g(x)] - \sum_{x \in E} g(x).$$

Выберем для элементов множества E произвольные значения функций $g(x)$ с тем единственным условием, что

$$\sum_{x \in E} g(x) = f(\rho_0),$$

и положим для любого X

$$g(X) = f(\rho_X) - f(\rho_0) - \sum_{x \in X} f(x).$$

Тем самым теорема доказана.

Замечание. Предположим, что на самом деле ρ_0 оказалось не самым „тонким" отношением эквивалентности, а существует еще более „тонкое" отношение ρ_0', отличающееся от ρ_0 только тем, что в ρ_0' класс (X) разбит на (X') и (X''). Каждое отношение эквивалентности ρ_i', менее „тонкое" чем ρ_0', всегда или $> \rho_0$, или имеет вид $(X')(X'')(U)(V)\ldots(W)$, т. е. совпадает на $E - X$ с некоторым $\rho_i > \rho_0$.

Следовательно, в этом случае

$$(\rho_0', \rho_i') \sim (\rho_0, \rho_i) \quad \text{и} \quad f(\rho_i') = f(\rho_i) + f(\rho_0') - f(\rho_0),$$

и $f(\rho_i')$ принимает искомый вид, если выбрав произвольное $g(X')$, положить

$$g(X'') = f(\rho_0') - f(\rho_0) - g(X') + g(X).$$

Предположим теперь, что из $\rho' < \rho$ следует $f(\rho) \leqslant f(\rho') \leqslant f < \infty$, и пусть можно найти такое конечное ρ_0, что $f - f(\rho_0) \leqslant \varepsilon$ и для всех классов Y отношения ρ_0

$$0 < g(Y) \leqslant \eta$$

Тогда такое же условие можно распространить на ρ_0', взяв, например,

$$0 < g(X'') = g(X') = \frac{1}{2}\left[f'(\rho_0') - f(\rho_0) + g(X)\right] < \frac{\varepsilon + \eta}{2}.$$

Это замечание приводит к возможности распространения теоремы I на любое E при соответствующем определении того, что понимается под отношением эквивалентности с бесконечным множеством классов, и при соответствующих ограничениях на $f(\cdot)$.

Возвращаясь к основной цели — измерению информации, мы постулируем следующее условие.

Условие I. *Мера информации $H(\rho_i)$, отнесенной к наблюдателю, является нормировкой на множестве (структуре) R всех отношений эквивалентности на E.*

Точное значение этого условия представляет довольно сильное требование. Именно, если Ω_i отличается от Ω_j такой же способностью к более тонким различениям, как Ω_k от Ω_l, то разности $H(\rho_i) - H(\rho_j)$ и $H(\rho_k) - H(\rho_l)$ должны быть равны. Если угодно, условие I можно истолковать также, уподобляя H стоимости оборудования и требуя, чтобы дополнительные расходы на вспомогательные приспособления, предназначенные для выполнения более тонкого анализа, не зависели от общей стоимости других частей наблюдательного механизма.

Из теоремы I следует только, что если число состояний конечно, то $H(\rho)$ есть сумма членов, зависящих только от классов отношения ρ. Конечно, столь широкое определение не позволяет задать H действительно интересным образом, и мы выставим дальнейшие постулаты.

Условие II. *Если $\rho < \rho'$ и притом для класса V отношения ρ' $\rho \equiv \rho'(E - V)$ (т. е. если Ω отличается от Ω' только дальнейшим разбиением одного из классов), то*

$$H(\rho) = H(\rho') + P(V)\, H(\rho''),$$

где $H(\rho'')$ — мера информации, отнесенной к $\rho'' = \rho\,[V]$ для наблюдателя, которому известно, что состояние ξ принадлежит к подмножеству V.

К этому условию мы присоединяем:

Условие III. *Вероятности $P(X) = x \neq 0$ суть элементы некоторого топологического коммутативного кольца $\mathfrak{A}$* [1]*, а $H(\rho)$ — непрерывный функционал от $x, y, \ldots$*

Условие IV. *Кольцо $\mathfrak{A}$ таково, что, если h_i есть непрерывный функционал, удовлетворяющий для всех $a, b \in \mathfrak{A}$ ($a + b = 1$, $ab \neq 0$) условиям:*

---

[1]) Грубо говоря, топологическое коммутативное кольцо есть множество, в котором абстрактные операции $+$ и $\times$ определены так, чтобы удовлетворялись обычные условия, за исключением равенства $ab = 0$, которое здесь может оказаться справедливым даже при $a \neq 0$ и $b \neq 0$.

1) $h_1(a+b)=h_1(a)+h_1(b)$, *то* $h_1(\cdot)=\Delta(\cdot)$,
где $\Delta(\cdot)$ — *линейный функционал;*

2) $h_2(ab)=h_2(a)+h_2'(b)$, *то* $h_2(\cdot)=\Delta\log(\cdot)$,
где $\Delta(\cdot)$ — *линейный функционал*, а $\log(\cdot)$ — *некоторая заданная функция внутри кольца функций над* $\mathfrak{A}$.

Теорема II. *При условии IV необходимым и достаточным условием того, чтобы* $H(\rho)$ *удовлетворяло условиям I, II и III, является равенство*

$$H(\rho)=\sum_x x\,\Delta\log(x),$$

где суммирование производится по классам X *отношения* ρ, *а* Δ *есть любой непрерывный линейный функционал.*

Доказательство. Достаточность является предметом непосредственной проверки. Что же касается необходимости, то условие I влечет

$$H(\rho)=\sum g'(x)=\sum xg(x)$$

для некоторого $g(X)=\frac{1}{x}g'(x)$.

Рассмотрим отношение эквивалентности с четырьмя классами $\rho(X)(Y)(Z)(T)$.

Из условия II при $V=(Y\dot{+}Z\dot{+}T)$ следует

$$H(\rho)=xg(x)+yg(y)+zg(z)+tg(t)=xg(x)+$$
$$+(y+z+t)\,g(y+z+t)+(y+z+t)\left[\frac{y}{y+z+t}g\left(\frac{y}{y+z+t}\right)+\right.$$
$$\left.+\frac{z}{y+z+t}g\left(\frac{z}{y+z+t}\right)+\frac{t}{y+z+t}g\left(\frac{t}{y+z+t}\right)\right].$$

Введем обозначения $y+z+t=s$, $y'=ys^{-1}$, $z'=zs^{-1}$, $t'=ts^{-1}$ и $k(a;\ b)=g(ab)-g(a)-g(b)$. Тогда после перегруппировки членов получим

$$yk(y';\ s)+zk(z';\ s)+tk(t';\ s)=0.$$

В частности, при $t=0$ и тех же значениях x и y

$$yk(y';\ s)+(z+t)\,k(z'+t';\ s)=0.$$

Из двух последних уравнений вычитанием получим

$$(z'+t')\,k(z'+t';\ s)=z'k(z';\ s)+t'k(t';\ s).$$

Следовательно, так как $ak(a,b)$ аддитивно по первому аргументу, то, в силу условия IV (1), $ak(a;b)=\Delta(a)$, где Δ может зависеть от b, но не от a. Однако $k(a;b)$ симметрично по a и b.

Это в свою очередь влечет $k(a; b) = \frac{1}{ab}\Delta_1(ba)$, и приведенное выше уравнение дает

$$\Delta_1(y's) + \Delta_1(z' + t')s = \Delta_1(y) + \Delta_1(z + t),$$

где y и $z + t$ ограничены единственным условием $y + z + t < 1$. Так как Δ_1 аддитивно, то отсюда вытекает $\Delta_1(u) = 0$ для всех $0 < u < 1$, так что

$$k(a; b) = g(ab) - g(a) - g(b) = 0$$

для всех $a, b \in \mathfrak{A}$, $a + b < 1$, $ab \neq 0$.

Теперь ввиду условия IV(2) получим $g(a) = \Delta \log a$, что завершает доказательство.

Замечание 1. Два последовательных условия I и II, содержащиеся в приведенной выше аксиоматике, можно заменить одним постулатом I′ (при выполнении условий III и IV и при конечном E).

Условие I′. Для всех $\rho = (X)(Y)(Z)$, $\rho_1' = (X)(Z \dot{+} Y)$, $\rho_1'' = \rho[E - X] = (Y)(Z)$, $\rho_2' = (X \dot{+} Y)(Z)$, $\rho_2'' = (X)(Z)$ справедливо соотношение $H(\rho_1') + P(X)H(\rho_1'') = H(\rho_2') + P(Z)H(\rho_2'')$.

В самом деле, можно показать, что условие I′ включает условие I. Эта формулировка, которая не содержит понятия нормировки, может быть истолкована как „принцип виртуального разложения наблюдений в последовательные дихотомии“, так как она требует, чтобы информация, связанная с различением между принадлежностью ξ к X, Y, Z, могла быть вычислена только лишь на основании информации, связанной с дихотомиями ρ' и ρ''.

Замечание 2. Если ограничиться информацией, зависящей только от численных значений $P(X)$, как в случае теории связи, то условие II можно заменить аддитивностью для сочетания независимых переменных (Вудворд). Тогда, *если оценка непрерывна и зависит только от $P(X)$, необходимым и достаточным условием ее аддитивности* для сочетания независимых переменных является ее представимость в виде

$$\sum x \log x.$$

Доказательство. Пусть η и ζ суть две независимые переменные, принимающие соответственно значения $Y_1Y_2Y_3$ и $Z_1Z_2Z_3$, и пусть $\xi = \eta \cdot \zeta$ есть их абстрактное произведение, принимающее 3×3 значений X_{ij} $(i, j = 1, 2, 3)$. Пусть η^* (соответственно ζ^*) есть переменная, полученная из η (соответственно из ζ) путем смешения значений № 2 и № 3. Так как предполагается, что H является нормировкой, то информация $H(\eta \cdot \zeta)$ о величине ξ является суммой $H(\eta \cdot \zeta) = \sum_{ij} x_{ij}\, g(x_{ij})$, которая по предположению равна

$$H(\eta) + H(\zeta) = \sum_i y_i g(y_i) + \sum_j z_j g(z_j).$$

Далее,

$$D = H(\eta \cdot \zeta) - H(\eta \cdot \zeta^*) - H(\eta^* \cdot \zeta) + H(\eta^* \cdot \zeta^*) =$$
$$= H(\eta) + H(\zeta) - [H(\eta) + H(\zeta^*)] - [H(\eta^*) +$$
$$+ H(\zeta)] + H(\eta^*) + H(\zeta^*) = 0.$$

С другой стороны, положив $x_{ij} = P(X_{ij})$, получим

$$D = x_{22}g(x_{22}) + x_{23}g(x_{23}) + x_{32}g(x_{32}) + x_{33}g(x_{33}) -$$
$$- [(x_{22} + x_{23})\, g(x_{22} + x_{23}) + (x_{32} + x_{33})\, g(x_{32} + x_{33}) +$$
$$+ (x_{22} + x_{32})\, g(x_{22} + x_{32}) + (x_{23} + x_{33})\, g(x_{23} + x_{33})] +$$
$$+ (x_{22} + x_{23} + x_{32} + x_{33})\, g(x_{22} + x_{23} + x_{32} + x_{33}) = 0.$$

Предположим теперь, что $z_2 = z_3 = z$. После перегруппировки членов последнее уравнение дает

$$y_2[2g(y_2z) - 2g(2y_2z)] + y_3[2g(y_3z) - 2g(2y_3z)] -$$
$$- (y_2 + y_3)\{2g[(y_2 + y_3)z] - 2g[2(y_2 + y_3)z]\} = 0.$$

Воспользовавшись обозначением $g(ab) - g(2ab) = k(a;\, b)$, получим равенство

$$y_2k(y_2;\, z) + y_3k(y_3;\, z) = (y_2 + y_3)\, k(y_2 + y_3;\, z),$$

из которого следует (так как непрерывность постулирована), что $k(y_i;\, z)$ не зависит от y_i.

Так как $k(y_i;\, z)$ симметрична по y_i и z, то она является постоянной K. Положив $u = y_iz$, окончательно получим

$$g(u) - g(2u) = K \quad \text{для всех} \quad 0 \leqslant u < 1.$$

Отсюда следует справедливость теоремы, так как мы получили уравнение Шрёдера, имеющее, как известно [7], единственное решение

$$g(u) = K \log u.$$

Заметим, что доказательство не удалось бы, если бы мы не предположили, что g и k суть числовые функции, так как нельзя было бы доказать, что $k(a;\, b)$ является постоянной. Действительно, в более общем случае не только информация, но и результаты применения линейных операций к ней (т. е. выражения вида $\sum \Delta_1 (x\, \Delta_i \log x)$) удовлетворяют требованию аддитивности для сочетания независимых переменных.

III

Теперь, когда мы получили общее выражение для меры информации, необходимо связать оператор Δ с тем, к чему относится соответствующая информация и что пока еще не было нами определено. Мы не намного продвинулись в этом направлении, но приводимые ниже собрамения могут послужить намеками на полный ответ.

Теорема. *Если имеется такая функция* $\tau(\xi)=\zeta$, *что для всех* ξ (*условные*) *значения* $P(\xi/\zeta)\Delta\log P(\xi/\zeta)$ *равны нулю, то* Δ-*информации о* ξ *и о* ζ *равны.*

Доказательство. Достаточно убедиться, что формула для условных информаций справедлива в общем случае, т. е. если ξ и η — две переменные, а $H(\xi\cdot\eta)$ — информация о ξ и η, то

$$H(\xi\cdot\eta)=H(\eta)+E_{\eta}H(\xi/\eta)$$

(где правая часть есть сумма информации о η и математического ожидания условной информации о ξ при известном η). Если теперь $\zeta=\eta$ функционально связана с ξ, то

$$H(\xi\cdot\zeta)=H(\xi)=H(\zeta),$$

так как, по предположению, $H(\xi/\zeta)=0$.

Например, пусть Δ есть оператор $[\ \ ]_{\theta=\theta_1}^{\theta=\theta_0}$ (значение при $\theta=\theta_0$ минус значение при $\theta=\theta_1$, где θ — неизвестный параметр, от которого зависит $P(\xi)$). С его помощью описывается информация W. Предположим, что распределение ξ можно представить в виде

$$P(\xi=\xi_0)=P(\zeta=\zeta_0)\,P(\zeta'=\zeta_0'),$$

где ζ — функция от ξ, а $P(\zeta'=\zeta_0)$ имеет одинаковые значения для $\theta=\theta_0$ и $\theta=\theta_1$. Хорошо известно, что в этом случае ζ представляет *„достаточную статистику"* для проверки $\theta=\theta_0$ или $\theta=\theta_1$.

Информация *T*. Следующий пример будет рассмотрен несколько подробнее, так как он является единственным примером подлинной „информации", если не считать классических H, F и W.

Пусть дано множество переключателей $a_1a_2\ldots a_i\ldots\ i\in I$, которые могут быть разомкнуты ($a_i=0$) или замкнуты ($a_i=1$) независимо друг от друга, причем вероятность разомкнутого состояния одинакова для всех и равна ω.

Рассмотрим множество произведений

$$A_1=\prod_{i\in I_1}a_i,\qquad A_2=\prod_{i\in I_2}a_i,\ \ldots,\ A_k=\prod_{i\in I_k}a_i,$$

где I_j суть заданные подмножества множества I. Наблюдение заключается в проверке одновременного выполнения равенств $A_1=A_2=\ldots=A_k=0$.

Нужно определить, для какого числа (ν) переключателей можно *доказать* с помощью такого наблюдения, что они в среднем разомкнуты. Конечно, если мы знаем только, что *не все* A_j равны нулю, то ничего *доказать* нельзя. Обозначим вероятность этого случая через Q.

Наоборот, если *все* $A_j = 0$ (с априорной вероятностью $P = 1 - Q$), то по крайней мере некоторые из A_i должны быть разомкнуты, причем их минимальное количество зависит от комбинаций, в которых a_i входят в множество произведений A_j. Например, если $A_1 = a_1a_2a_3$, $A_2 = a_1a_4a_5$, $A_3 = a_2a_4$, то ν равно двум, что соответствует трем возможностям:

$$a_1 = a_2 = 0; \qquad a_1 = a_4 = 0; \qquad a_2 = a_4 = 0.$$

Введем оператор Δ^* — произведение ω на частную производную по ω при $\omega = 0$:

$$\Delta^* = \left[\omega \frac{\partial}{\partial\omega}\right]_{\omega=0}$$

Соответствующая информация равна

$$T = P\Delta^* \log P + Q\Delta^* \log Q.$$

Так как событие вероятности Q может случиться даже, когда все a_i замкнуты, то Q имеет вид $1 - \omega^\alpha R$, где R — многочлен с ненулевым постоянным членом, так что

$$\Delta^* \log Q = \left[\frac{-\omega(\omega^\alpha R' + \alpha\omega^{\alpha-1}R)}{1 - \omega^\alpha R}\right]_{\omega=0} = 0.$$

С другой стороны, $P = \omega^\alpha R$, где α в точности равно тому минимальному числу переключателей (ν), которые должны быть разомкнуты для того, чтобы удовлетворялось равенство $A_1 = A_2 = \ldots = A_k = 0$.

В этом легко убедиться прямым вычислением. Тогда

$$\Delta^* \log P = \left[\omega\left(\frac{\alpha}{\omega} + \frac{R'}{R}\right)\right]_{\omega=0} = \alpha = \nu,$$

а T в точности равно среднему значению νP этого минимального числа.

С познавательной точки зрения интересно отметить, что T представляет собой промежуточное понятие между случаем отсутствия шумов, когда H служит рациональным инструментом, и случаем наличия шумов, когда ведущую роль играет W или F, в соответствии с функцией стоимости, которая задается внутренней топологией „сообщений“ (последние при наличии шума суть параметры θ, а не ξ).

В самом деле, T введено в „полудетерминистическом случае“, когда только некоторые сообщения (в данном случае сообщение „все $A_j = 0$“) допускают абсолютно безусловную интерпретацию.

ЛИТЕРАТУРА

1. Bartlett M. S., *Proc. London Symp. Inf. Theory*, p. 81 (1950).
2. Darmois G., *Rev. Inst. Int.* Stat. (1945).
3. Fisher R. A., *Phil. Trans. Roy. Soc.*, A 22, 309 (1921).
4. Frechet M., *Rev. Inst. Int. Stat.*, (3/4) p. 182 (1942).
5. Kullback, *Ann. Math. Stat.*, 2, 745 (1954).
6. Rao G. Q., *Bull. Calcutta Math. Soc.*, 37, 81 (1945).
7. Schröder, *Math. Ann.*, 3, 296 (1871).
8. Schützenberger M. P., *Pub. Inst. Stat. Univ. Paris*, 3, 27 (1953).
9. Wald A., Sequential Analysis, Wiley and Son, N. Y. (1947).
10. Woodward, *Proc. IEE* (1952).