ИДЕНТИФИКАЦИЯ СТАТИСТИЧЕСКИХ ПАРАМЕТРОВ
В ОДНОЙ МОДЕЛИ УСЛОВНОЙ НЕЗАВИСИМОСТИ

М.И.Шлезингер

1. Описание модели

Пусть некоторый объект характеризуется тремя параметрами: x, y и k, где x, y и k принимают значения из некоторых конечных множеств X, Y и K, причем множество K состоит всего из двух значений: 1 и 2. Параметры x и y являются наблюдаемыми параметрами и называются в дальнейшем признаками объекта. Параметр k является ненаблюдаемым параметром и называется состоянием объекта. Тройка параметров x, y и k является случайной и принимает те или иные значения в соответствии с распределением вероятностей $p^*: X \times Y \times K \to R$. Для любых $x \in X,\ y \in Y,\ k \in K$ число $p(x,y,k)$ обозначает вероятность того, что объект находится в состоянии k и характеризуется признаками x и y.

Вероятности $p^*(x,y,k),\ x \in X,\ y \in Y,\ k \in K$ однозначно определяют маргинальные и условные вероятности вида

$$p^*_{XY}(x,y) = \sum_{k \in K} p^*(x,y,k) \;;$$

$$p^*_X(x) = \sum_{y \in Y} p^*_{XY}(x,y) \;;$$

$$p^*_Y(y) = \sum_{x \in X} p^*_{XY}(x,y) \;;$$

$$p^*_K(k) = \sum_{x \in X} \sum_{y \in Y} p^*(x,y,k) \;;$$

$$p^*_{X,Y/K}(x,y/k) = \frac{p^*(x,y,k)}{p^*_K(k)} \;;$$

$$p^*_{X/Y}(x/y) = \frac{p^*_{X,Y}(x,y)}{p^*_Y(y)} \;;$$

$$p^*_{Y/X}(y/x) = \frac{p^*_{X,Y}(x,y)}{p^*_X(x)}$$

и аналогично другие маргинальные и условные вероятности.

В дальнейшем, при использовании приведенных обозначений, нижний индекс в идентификаторе функции, как правило, будет совпадать со списком аргументов внутри скобок. Поэтому в тех случаях, где это не будет приводить к неоднозначному пониманию, этот нижний индекс в идентификаторе функции будет опускаться. Так, например, обозначение $p^*(x,y\ /\ k)$ будет использоваться вместо $p^*_{X,Y/K}(x,y\ /\ k)$ и т.п.

Совокупность вероятностей $p^*(x,y,k),\ x \in X,\ y \in Y,\ k \in K$, будет называться статистической моделью исследуемого объекта. В данной работе эта модель исследуется в том случае, когда она является моделью условной независимости, а именно, когда при каждом фиксированном ненаблюдаемом состоянии k наблюдаемые признаки взаимно независимы. Формально говоря, нами будут исследованы модели, для которых

$$p^*(x,y\ /\ k)\ =\ p^*(x\ /\ k)\ \cdot\ p^*(y\ /\ k)\ . \tag{1}$$

Условная независимость признаков x и y при любом фиксированном состоянии k ни в коем случае не обозначает их безусловной независимости. Это значит, что из (1) никоим образом не следует, что

$$p^*(x,y)\ =\ p^*(x)\ \cdot\ p^*(y)\ . \tag{2}$$

Условная независимость наблюдаемых признаков x и y объекта не отрицает их зависимости, а констатирует тот факт, что эта зависимость осуществляется через зависимость от ненаблюдаемого состояния k объекта.

Цель данной статьи состоит в том, чтобы ответить на вопрос об идентифицируемости модели объекта по наблюдению только тех ее параметров, которые являются наблюдаемыми. Мы хотим узнать, в какой мере можно по наблюдению потенциально бесконечной последовательности (x_1,y_1), (x_2,y_2), ..., (x_i,y_i), ... наблюдений восстановить не только совместные вероятности $p^*(x,y)$, но и совместные вероятности $p^*(x,y,k)$. Или, иными словами говоря, можно ли, не наблюдая никогда состояние k объекта, восстановить его распределение вероятностей $p^*(k)$ и указать, как от этого состояния зависят наблюдаемые признаки x и y.

Вторая не менее важная цель работы состоит в конструировании конкретного алгоритма, который эту идентификацию производит. Этот алгоритм сформулирован нами, как частный случай ранее разработанного мною общего алгоритма идентификации смеси [1]. Этот алгоритм основан на сведении задачи идентификации смеси к решению вполне определенной оптимизационной задачи. В общем случае этот алгоритм обеспечивает сходимость к тому или иному локальному максимуму определенного функционала.

Важным новым результатом данной работы является установление того факта, что для рассматриваемого в данной работе класса моделей имеется в определенном смысле единственный локальный максимум, который, следовательно, является и глобальным.

Исследуемая в данной работе модель является, по-видимому, простейшей моделью условной независимости. Полученные результаты позволяют надеяться, что они могут быть обобщены и на случай более сложных моделей (в частности, на первом шаге для случая, когда $|K|>2$). Однако такое обобщение вряд ли может быть достигнуто без значительных дополнительных условий. Как можно будет увидеть по дальнейшему содержанию статьи, анализ одноэкстремальности даже для простейшей задачи не являются слишком простым.

2. Иденцифицируемость модели

Введем дополнительное предположение о рассматриваемой модели, которое будем называть предположением о существовании идеальных представителей. А именно, предполагается, что один из признаков, допустим, признак x таков, что существуют два его значения x_1 и x_2 такие, что

$$p^*(x_1 \ / k=1) \neq 0,\ p^*(x_1 \ / k=2)=0,\ p^*(x_2 \ / k=1)=0,\ p^*(x_2 \ / k=2)=0 \ . \tag{3}$$

Значение x_1 будем называть идеальным представителем первого состояния, а x_2 - идеальным представителем второго состояния. Данное предположение констатирует лишь существование таких представителей и не предполагает, что эти представители заранее известны. О признаке y будем предполагать лишь, что он зависит от состояния k, т.е., что

$$p^*(y \ / k=1) \ \neq \ p^*(y \ / k=2) \tag{4}$$

по крайней мере для одного значения $y \in Y$.

И наконец, вполне естественным является предположение, что

$$p^*(k=1) \neq 0 \quad \text{и} \quad p^*(k=2) \neq 0 \ . \tag{5}$$

При сделанных предположениях статистическая модель объекта полностью идентифицируема. На основании достаточно длинной последовательности $(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_i, y_i),\ \ldots$ могут быть восстановлены вероятности $p^*(x,y)$, а затем вычислены вероятности $p^*(x,y,k)$ такие, что для любой пары $(x,y) \in X \times Y$ справедливо равенство

$$p^*(x,y) = \sum_k p^*(x,y,k) \ . \tag{6}$$

Мы покажем, что существуют лишь две функции $p^*(x,y,k)$ удовлетворяющие условиям (1), (3)-(6), и эти две функции отличаются друг от друга лишь переименованием состояний k. Более строгую формулировку и обоснование этих утверждений дает следующая теорема и ее доказательство.

Теорема. Для любой совокупности чисел $p^*(x,y),\ x \in X,\ y \in Y$, существует не более двух различных совокупностей $p^*(x,y,k),\ x \in X,\ y \in Y,\ k \in K$, удовлетворяющих условиям

$$p^*(x,y) = \sum_k p^*(x,y,k) \ , \tag{7}$$

$$\exists\, x \in X \Big(p^*(x \ / k=1) \neq 0 \ \&\ p^*(x \ / k=2) = 0 \Big) \ , \tag{8}$$

$$\exists\, x \in X \Big(p^*(x \ / k=1) = 0 \ \&\ p^*(x \ / k=2) \neq 0 \Big) \ , \tag{9}$$

$$\exists\, y \in Y \Big(p^*(y \ / k=1) \neq \ p^*(y \ / k=2) \Big) \ , \tag{10}$$

$$p^*(k=1) \neq 0,\ p^*(k=2) \neq 0 \ , \tag{11}$$

$$\forall (x,y,k) \in X \times Y \times K \Big(p^*(x,y,k) = p^*(k) \cdot p^*(x \ / k) \cdot p^*(y \ / k) \Big) \ . \tag{12}$$

При этом, если $p_1^*(x,y,k)$ и $p_2^*(x,y,k)$, $x \in X$, $y \in Y$, $k \in K$, - две совокупности чисел, удовлетворяющие (7-12), то для этих двух совокупностей справедливо

$$\forall x, y \ \left(p_1^*(x,y,k=1) = p_2^*(x,y,k=2)\right) \ , \tag{13}$$

$$\forall x, y \ \left(p_2^*(x,y,k=2) = p_1^*(x,y,k=1)\right) \ . \tag{14}$$

Доказательство. Мы докажем, что, если существует совокупность чисел $p^*(x,y,k)$, $x \in X$, $y \in Y$, $k \in K$, удовлетворяющая условиям (7-12), то на основании совокупности чисел $p^*(x,y)$, $x \in X$, $y \in Y$, единственным образом восстанавливается пара функций, одна из которых (неизвестно какая) равна $p^*(y \ / \ k=1)$, а вторая равна $p^*(y \ / \ k=2)$.

Представим совокупность чисел $p^*(x,y)$, $x \in X$, $y \in Y$, в виде таблицы, состоящей из $|X|$ столбцов и $|Y|$ строк. Строки этой таблицы соответствуют различным значениям признака y , столбцы - различным значениям признака y , а в пересечении x -ого столбца и y -ой строки записана вероятность $p^*(x,y)$.

Рассматривая столбцы этой таблицы, как некоторые $|Y|$ - мерные векторы, построим линейное замыкание этой совокупности векторов. Это линейное замыкание образует двумерное линейное пространство. Действительно, для каждого $x \in X$ вектор, записанный в x -ом столбце, т.е. вектор, компоненты которого равны $p^*(x,y), y \in Y$, является линейной комбинацией двух $|Y|$ -мерных векторов с компонентами $p^*(y \ / \ k=1)$, $y \in Y$, и $p^*(y \ / \ k=2)$, $y \in Y$, соответственно, т.к. в силу условия (7) и (12) для любого $x \in X$ справедливо

$$p^*(x,y) = p^*(k=1,x) \cdot p^*(y \ / \ k=1) + p^*(k=2,x) \cdot p^*(y \ / \ k=2) \ . \tag{15}$$

Векторы же $p^*(y \ / \ k=1)$ и $p^*(y \ / \ k=2)$ не являются линейно зависимыми, т.к. оба они ненулевые и по условию (10) не равны друг другу, а следовательно, неколинеарны.

Среди столбцов построенной нами таблицы имеются столбцы, соответствующие идеальным представителям, существование которых гарантируется по условиям (8) и (9). Эти столбцы могут быть обнаружены в соответствии со следующими правилами.

1. Исключим из таблицы те столбцы, в которых записаны нулевые векторы. Это те столбцы, которые соответствуют тем значениям x , для которых $p^*(x)=0$. Эти столбцы не соответствуют идеальным представителям. Действительно, вероятность $p^*(x)$ для идеального представителя не может равняться нулю, т.к. $p^*(x) = p^*(k=1) \cdot p^*(x \ / \ k=1) + p^*(k=2) \cdot p^*(x \ / \ k=2)$. В этом выражении величины $p^*(k=1)$ и $p^*(k=2)$ не равны нулю по условию (11), а одна из величин $p^*(x / k=1)$ и $p^*(x / k=2)$ не равна нулю по условию (8) и (9).

2. Исключим из оставшейся таблицы столбцы, соответствующие таким значениям x, для которых существуют такие значения x_1 и x_2 и такие числа $\alpha_1 > 0$ и $\alpha_2 > 0$, что равенство

$$p^*(x,y) = \alpha_1 \cdot p^*(x_1,y) + \alpha_2 \cdot p^*(x_2,y) \tag{16}$$

выполняется для всех $y \in Y$, а равенство

$$p^*(x_1,y) = p^*(x_2,y) \tag{17}$$

выполняется не для всех $y \in Y$. Докажем, что эти столбцы не соответствуют идеальным представителям.

С учетом (15) условие (16) может быть преобразовано к виду

$$\begin{aligned} p^*(x,y) &= \alpha_1 \cdot p^*(x_1,y) + \alpha_2 \cdot p^*(x_2,y) = \\ &= \left(\alpha_1 \cdot p^*(k=1,x_1) + \alpha_2 \cdot p^*(k=1,x_2)\right) \cdot p^*(y \,/\, k=1) + \\ &+ \left(\alpha_1 \cdot p^*(k=2,x_1) + \alpha_2 \cdot p^*(k=2,x_2)\right) \cdot p^*(y \,/\, k=2) \,. \end{aligned}$$

Из этого равенства следует, что

$$p^*(k=1,x) = \alpha_1 \cdot p^*(k=1,x_1) + \alpha_2 \cdot p^*(k=1,x_2) \tag{18}$$

и

$$p^*(k=2,x) = \alpha_1 \cdot p^*(k=2,x_1) + \alpha_2 \cdot p^*(k=2,x_2) \,. \tag{19}$$

Это в свою очередь обозначает, что $p^*(k=1,x) \neq 0$, т.к. $\alpha_1 > 0$, $\alpha_2 > 0$, а вероятности $p^*(k=1,x_1)$ и $p^*(k=1,x_2)$ не могут обе одновременно равняться нулю, т.к. в силу (17) они не равны друг другу. Однако для идеального представителя одна из вероятностей $p^*(k=1,x)$ и $p^*(k=2,x)$ обязана равняться нулю и поэтому рассматриваемое значение x не является идеальным представителем.

После выполненных исключений в таблице останутся только столбцы, соответствующие тем значениям x, для которых либо $p^*(k=1,x) = 0$, а $p^*(k=2,x) \neq 0$, либо $p^*(k=1,x) \neq 0$, а $p^*(k=2,x) = 0$, т.е. соответствующие одним лишь идеальным представителям. Вектор $p^*(x,y),\ y \in Y$, записанный в каждом из оставшихся столбцов, равен либо $p^*(k=1,x) \cdot p^*(y \,/\, k=1)$, либо $p^*(k=2,x) \cdot p^*(y \,/\, k=2)$. Следовательно, функция $\dfrac{p^*(x,y)}{\sum_{y \in Y} p^*(x,y)}$, вычисленная для каждого неисключенного значения x, есть либо $p^*(y \,/\, k=1)$, либо $p^*(y \,/\, k=2)$.

Таким образом, на основании вероятностей $p^*(x,y)$ единственным образом может быть построена пара функций, одна из которых (неизвестно какая) равна $p^*(y\ /\ k=1)$, а вторая - $p^*(y\ /\ k=2)$. Это значит, что функция $p^*(y\ /\ k)$ может быть восстановлена с точностью до перестановки значений переменной k .

Для каждого из этих двух вариантов функции $p^*(y\ /\ k)$ единственным образом восстанавливается функция $p^*(x,k)$. Это следует из того, что для каждого значения x функция $p^*(x,k)$ может быть единственным образом представлена в виде (15). Вероятности же $p^*(x,y,k)$ вычисляются с помощью простого умножения вероятностей $p^*(x,k)$ и $p^*(y/k)$.

Теорема доказана.

Выполненные рассуждения, доказывающие идентифицируемость модели условной независимости в случае, когда ненаблюдаемый параметр k принимает лишь два значения, могут быть, конечно же, обобщены и для случая, когда этот параметр принимает произвольное, но заранее известное количество значений. В следующем разделе будут описаны и обоснованы пути построения алгоритмов идентификации. К сожалению, это обоснование выполнено лишь для случая $K=\{1,2\}$ и в настоящее время не найдены какие-либо очевидные пути обобщения этих результатов на случай произвольного количества значений ненаблюдаемого параметра.

3. Формулировка задачи идентификации модели и пути ее решения

Идентификацию модели объекта по последовательности параметров этого объекта естественно рассматривать, как частный случай идентификации компонент смеси по выборке из этой смеси. Действительно, в нашей модели речь идет о двух генеральных совокупностях. Первая из них - это совокупность пар значений x,y , наблюдаемых при условии, что объект находится в первом состоянии. Распределение вероятностей в этой совокупности имеет вид $p(x\ /\ k=1)\cdot p(y\ /\ k=1)$. Вторая генеральная совокупность - это совокупность пар значений, наблюдаемых при условии , что объект находится во втором состоянии. В этой совокупности распределение вероятностей имеет вид $p(x\ /\ k=2)\cdot p(y\ /\ k=2)$. Из этих двух совокупностей образована смесь с весами $p(k=1)$ и $p(k=2)$ и получена, таким образом, новая совокупность, имеющая распределение $p(k=1)\cdot\ p(x\ /\ k=1)\cdot p(y\ /\ k=1)+p(k=2)\cdot\ p(x\ /\ k=2)\cdot p(y\ /\ k=2)$. Из этой смешанной совокупности взята выборка наблюдений $(x_1,y_1),\ (x_2,y_2),\ \ldots,\ (x_m,y_m)$. На основании этой выборки должна быть сделана оценка распределений $p(x\ /\ k=1),\ p(y/\ k=1),\ p(x\ /\ k=2),\ p(y\ /\ k=2)$, характеризующих компоненты смеси, и вероятностей $p(k=1)$ и $p(k=2)$, определяющих, с какими весами компоненты присутствуют в смеси. Эту задачу естественно формулировать, как отыскание таких вероятностей $p(x\ /\ k),\ p(y\ /\ k),\ p(k),\ k=1{,}2$, которые максимизируют вероятность заданной последовательности наблюдений $(x_1,y_1),\ (x_2,y_2),\ \ldots,\ (x_n,y_n)$. Это значит, что задача сводится к максимизации функционала

$$F = \sum_i \log \sum_k p(k) \cdot p(x_i \ / k) \cdot p(y_i \ / k) \ . \qquad (20)$$

В более общем виде эта задача была сформулирована в работе [1]. Там же был сформулирован общий алгоритм ее решения. Применительно к рассматриваемому частному случаю этот алгоритм имеет следующий вид.

Обозначим через $p^*(x,y)$ число $\dfrac{m(x,y)}{m}$, где $m(x,y)$ указывает, сколько раз пара (x,y) встретилась в выборке наблюдений $(x_1,y_1), (x_2,y_2), \ldots, (x_m,y_m)$. Очевидно, что функционал (20) может быть записан в виде

$$F = \sum_{x,y} p^*(x,y) \log \sum_k p_k \cdot p(x \ / k) \cdot p(y \ / k) \ . \qquad (21)$$

Совокупность варьируемых величин $(p(k), p(x \ / k), p(y \ / k), \ x \in X, \ y \in Y, \ k \in K)$ обозначим через A. Алгоритм начинает работу с произвольного значения A^0, а затем шаг за шагом строит последовательность $A^0, A^1, A^2, \ldots, A^t, \ldots$ и т.д. Для получения A^{t+1} по ранее полученному A^t должны быть выполнены следующие вычисления.

$$\alpha_{\mathrm{k}}(x,y) = \frac{p^t(k) \cdot p^t(x \ / k) \cdot p^t(y \ / k)}{\sum_{k \in K} p^t(k) \cdot p^t(x \ / k) \cdot p^t(y \ / k)}, \ k \in K, \ \ x,y \in X \times Y \ ; \qquad (22)$$

$$p^{t+1}(k) \ = \sum_{x,y} \alpha_{\mathrm{k}}(x,y) \cdot p^*(x,y), \quad k \in K \ ; \qquad (23)$$

$$p^{t+1}(x / k) \ = \ \frac{\sum_y a_k(x,y) \cdot p^*(x,y)}{p^{t+1}(k)}, \quad k \in K \, , \ x \in X \, ; \qquad (24)$$

$$p^{t+1}(y / k) \ = \ \frac{\sum_x a_k(x,y) \cdot p^*(x,y)}{p^{t+1}(k)}, \quad k \in K \, , \ y \in Y \, . \qquad (25)$$

Указанную совокупность вычислений, преобразующих A^t в A^{t+1}, обозначим через T, так что $T\left(A^t\right) = A^{t+1}$, а значения A, удовлетворяющие равенству $A = T(A)$, будем называть неподвижными точками алгоритма.

В [2] было доказано, что при любом начальном значении A последовательность $A, \ T(A), \ T^2(A), \ldots, \ T^t(A), \ldots$ имеет предел, и этим пределом является неподвижная точка алгоритма. Мы докажем здесь, что условие неподвижности точки совпадает с хорошо известными классическими необходимыми условиями максимума функционала (20) и (21). Затем будет доказано, что при определенных условиях для исследуемой нами модели эти классические необходимые условия являются и достаточными.

Выражение (21) есть переменная величина, зависящая от $(|X| + |Y| + 1) \cdot |K|$ переменных $(p(k), \ k \in K), \ (p(x \ / k), \ x \in X, k \in K), \ (p(y \ / k), \ y \in Y, k \in K)$, которая должна максимизироваться при следующих $2 \cdot |K| + 1$ условиях

$$\sum_{k\in K} p(k)=1\ ,$$
$$\sum_{x\in X} p(x\ /\ k)=1,\ k\in K\ ,$$
$$\sum_{y\in Y} p(y\ /\ k)=1,\ k\in K\ .$$

Такой задаче на условный экстремум соответствует функция Лагранжа

$$\Phi(A)=\sum_{x,y} p^*(x,y)\ \log \sum_{k} p(k)\cdot p(x\ /\ k)\cdot p(y\ /\ k)+$$
$$+\lambda\cdot \sum_{k\in K} p(k)+\sum_{k}\gamma_k\cdot \sum_{x\in X} p(x\ /\ k)+\sum_{k}\eta_k \sum_{y\in Y} p(y\ /\ k)\ ,$$

а необходимые условия оптимальности выражаются следующей системой уравнений:

$$\left.\begin{aligned}
&\frac{\partial\ \Phi(A)}{\partial\ p(k)}=\sum_{x,y}\frac{p^*(x,y)}{p(x,y)}\cdot p(x\ /\ k)\cdot p(y\ /\ k)+\lambda\ =0,\ \ k\in K\ , && \text{(a)}\\
&\frac{\partial\ \Phi(A)}{\partial\ p(x\ /\ k)}=\sum_{y}\frac{p^*(x,y)}{p(x,y)}\cdot p(k)\cdot p(y\ /\ k)+\gamma_k\ =0,\ \ x\in X,\ k\in K\ , && \text{(b)}\\
&\frac{\partial\ \Phi(A)}{\partial\ p(y\ /\ k)}=\sum_{x}\frac{p^*(x,y)}{p(x,y)}\cdot p(k)\cdot p(x\ /\ k)+\eta_k\ =0,\ \ y\in Y,\ k\in K\ , && \text{(c)}\\
&\sum_{k} p(k)=1\ , && \text{(d)}\\
&\sum_{x} p(x/k)=1,\ k\in K\ , && \text{(e)}\\
&\sum_{y} p(y\ /\ k)=1,\ k\in K\ , && \text{(f)}
\end{aligned}\right\}(26)$$

Соотношение между условием (26) и неподвижностью значения A выражается теоремой.

Теорема. Если для любого $k\in K\ \ p(k)\neq 0$, то условие $A=T(A)$ эквивалентно системе уравнений (26).

Доказательство. Используем выражения (22-25) и запишем их же, исключив из этих выражений промежуточные переменные $\alpha_k(x,y)$. Для этого подставим выражение (22) для $\alpha_k(x,y)$ в (23), (24), (25) и получим

$$p^{t+1}(k) = \sum_{x,y} \frac{p^*(x,y)}{p^t(x,y)} \cdot p^t(k) \cdot p^t(x / k) \cdot p^t(y / k), \quad k \in K \ ; \tag{27}$$

$$p^{t+1}(x / k) = \frac{\sum_y \frac{p^*(x,y)}{p(x,y)} \cdot p^t(k) \cdot p^t(x / k) \cdot p^t(y / k)}{p^{t+1}(k)}, \quad k \in K,\ x \in X \ ; \tag{28}$$

$$p^{t+1}(y / k) = \frac{\sum_x \frac{p^*(x,y)}{p(x,y)} \cdot p^t(k) \cdot p^t(x / k) \cdot p^t(y / k)}{p^{t+1}(k)}, \quad k \in K,\ y \in Y \ . \tag{29}$$

Условие неподвижности точек получается, если в уравнениях (27), (28), (29) величины $p^t(k), p^t(x / k)$ и $p^t(y / k)$ приравнять к величинам $p^{t+1}(k), p^{t+1}(x / k), p^{t+1}(y / k)$, т.е. если в этих уравнениях просто опустить индексы t и $t+1$. После такого преобразования условия (27), (28), (29) приобретают вид

$$\sum_{x,y} \frac{p^*(x,y)}{p(x,y)} \cdot p(x / k) \cdot p(y / k) = 1, \ \ k \in K \ ; \tag{30}$$

$$\sum_y \frac{p^*(x,y)}{p(x,y)} \cdot p(y / k) = 1, \ \ k \in K,\ x \in X \ ; \tag{31}$$

$$\sum_x \frac{p^*(x,y)}{p(x,y)} \cdot p(x / k) = 1, \ \ k \in K,\ y \in Y \ . \tag{32}$$

Выясним, чему должны быть равны неопределенные коэфициенты Лагранжа $\lambda,\ \eta_k$ и γ_k в уравнениях (26,a,b,c) с учетом того, что должны выполняться условия (26, d,e,f).

Умножим уравнение (26,a) на $p(k)$, просуммируем это произведение по всем k и в результате получим, что

$$\sum_k \lambda \cdot p(k) = \lambda = -\sum_k p(k) \sum_{x,y} \frac{p^*(x,y)}{p(x,y)} \cdot p(x / k) \cdot p(y / k) =$$

$$= -\sum_{x,y} \frac{p^*(x,y)}{p(x,y)} \sum_k p(k) \cdot p(x / k) \cdot p(y / k) = -\sum_{x,y} p^*(x,y) = -1 \ ,$$

т.е. $\lambda = -1$, и условие (26) приобретает вид

$$\sum_{x,y} \frac{p^*(x,y)}{p(x,y)} \cdot p(x\ /\ k) \cdot p(y\ /\ k) = 1 \ , \tag{33}$$

в точности совпадающий с (30).

Умножим уравнение (26,b) на $p(x\ /\ k)$, просуммируем его по всем x и в результате получаем

$$\sum_x \gamma_k \cdot p(x\ /\ k) = \gamma_k = -\sum_x p(x\ /\ k) \sum_y \frac{p^*(x,y)}{p(x,y)} \cdot p(k) \cdot p(y\ /\ k) =$$

$$= -p(k) \sum_{x,y} \frac{p^*(x,y)}{p(x,y)} \cdot p(x\ /\ k) \cdot p(y\ /\ k) \ .$$

Учитывая (33) получаем, что $\gamma_k = p(k)$, и условие (26,b) становится эквивалентным условию (31).

Подобным образом, умножая равенство (26,c) на $p(y\ /\ k)$ и суммируя его по всем y, доказывается, что $y_k = -p(k)$ и что условие (26,c) эквивалентно (32).

Теорема доказана.

Мы докажем теперь, что неподвижность алгоритма является не только необходимым условием максимума функционала (21), но и в определенном смысле достаточным условием его глобального максимума. А именно, мы покажем, что условие неподвижности алгоритма можно усилить, то есть дополнить определенными легко распознаваемыми условиями так, что после этого усиления они становятся достаточными условиями глобального максимума. Более строго это утверждение формулируется следующей теоремой.

Теорема: Пусть функция $p^*(x,y)$ такова, что существуют такие числа $p^*(k)$ и функции $p^*(x/k),\ p^*(y/k),\ k=1,2$, что $p^*(x,y) = \sum_{\mathrm{k}} p^*(k) \cdot p^*(x/k) \cdot p^*(y/k)$; пусть также числа $p(k)$ и функции $p(x/k),\ p(y/k),\ k=1,2$, удовлетворяют уравнениям

$$\sum_x \frac{p^*(x,y)}{\sum_k p(k) \cdot p(x\ /\ k) \cdot p(y\ /\ k)} \cdot p(x\ /\ k) = 1 \quad , \tag{34}$$

для всех y и k и

$$\sum_y \frac{p^*(x,y)}{\sum_k p(k) \cdot p(x\ /\ k) \cdot p(y\ /\ k)} \cdot p(y\ /\ k) = 1 \tag{35}$$

для всех x и k;

в таком случае либо для любых x, y и k выполняются равенства $p(x\ /\ k) = p(x)$ и

$p(y\ /\ k)=p(y)$, либо для любых чисел $p'(k)$, и функций $p'(x\ /\ k),\ p'(y\ /\ k)$ выполняется неравенство

$$\sum_{x,y} p^*(x,y)\log\sum_k p(k)\cdot\ p(x/k)\ \cdot p(y/k) \geq \sum_{x,y} p^*(x,y)\log\sum_k p'(k)\cdot\ p'(x/k)\ \cdot p'(y/k)\ . \qquad (36)$$

Доказательство.

Естественно, доказательство теоремы сводится к доказательству того, что, если выполняется хотя бы одно из неравенств $p(x\ /\ k)\neq p(x)$ или $p(y\ /\ k)\neq p(y)$, то из условий теоремы следует неравенство (36). Мы выполним это доказательство для случая, когда $p(y\ /\ k)\neq p(y)$. Это доказательство представлено в виде следующих 12 доказанных утверждений.

1. Справедливость равенства (34) приводит к справедливости и следующих равенств:

$$\sum_k p(k)\ \cdot p(y\ /\ k)\sum_x \frac{p^*(x,y)\cdot p(x\ /\ k)}{\sum_k p(k)\cdot\ p(x\ /\ k)\ \cdot p(y\ /\ k)} = \sum_k p(k)\ \cdot p(y\ /\ k)\quad ;$$

$$\sum_x \frac{p^*(x,y)\cdot \sum_k p(k)\cdot\ p(x\ /\ k)\ \cdot p(y\ /\ k)}{\sum_k p(k)\cdot\ p(x\ /\ k)\ \cdot p(y\ /\ k)} = \sum_k p(y,k)\quad ;$$

$$\sum_x p^*(x,y)=\ p(y)\quad ;$$

$$p^*(y)=p(y)\ . \qquad (37)$$

Подобным образом доказывается, что из равенства (35) следует равенство

$$p^*(x)=p(x)\ . \qquad (38)$$

2. В силу (37) равенство (34) может быть записано в виде

$$\sum_x \frac{p^*(x\ /\ y)}{p(x\ /\ y)}\cdot p(x\ /\ k)=1\quad , \qquad (39)$$

а равенство (35) и силу (38) - в виде

$$\sum_y \frac{p^*(y\ /\ x)}{p(y\ /\ x)}\cdot p(y\ /\ k)=1,\ \ \forall k,\ \forall x\ ., \qquad (40)$$

3. Пусть x' - некоторое значение переменной x ; тогда для любого x существует такое число $\alpha(x)$, что равенство

$$p^*(y\ /\ x') - p^*(y)=\alpha(x)\cdot\ \left(p^*(y\ /\ x)-p^*(y)\right) \qquad (41)$$

выполняется для любого значения y, если только по крайней мере для некоторых значений y выполняется неравенство $p^*(y\ /\ x) \neq p^*(y)$. Действительно, справедливы равенства

$$p^*(y) = p^*(k=1)\cdot p^*(y\ /\ k=1) + p^*(k=2)\cdot p^*(y\ /\ k=2)\ ; \tag{42}$$

$$p^*(y\ /\ x) = p^*(k=1\ /\ x)\cdot p^*(y\ /\ k=1) + p^*(k=2\ /\ x)\cdot p^*(y\ /\ k=2)\ ; \tag{43}$$

$$p^*(y\ /\ x') = p^*(k=1\ /\ x')\cdot p^*(y\ /\ k=1) + p^*(k=2\ /\ x')\cdot p^*(y\ /\ k=2)\ . \tag{44}$$

Из (42) и (43) следует равенство

$$p^*(y\ /\ x) - p^*(y) = \left[p^*(k=1\ /\ x) - p^*(k=1)\right]\cdot\left[p^*(y\ /\ k=1) - p^*(y\ /\ k=2)\right]\ , \tag{45}$$

а из (42) и (44) - равенство

$$p^*(y\ /\ x') - p^*(y) = \left[p^*(k=1\ /\ x') - p^*(k=1)\right]\cdot\left[p^*(y\ /\ k=1) - p^*(y\ /\ k=2)\right]\ . \tag{46}$$

Поскольку неравенство $p^*(y\ /\ x) \neq p^*(y)$ выполняется по крайней мере для некоторых y, то $p^*(k=1\ /\ x) - p^*(k=1) \neq 0$ и справедливо выражение

$$p^*(y\ /\ x') \text{ - } p^*(y) = \frac{p^*(k=1\ /\ x') - p^*(k=1)}{p^*(k=1\ /\ x) - p^*(k)}\left[p^*(y\ /\ x) - p^*(y)\right]\ , \tag{47}$$

что доказывает (41).

4. Лемма 1. Пусть x_1 и x_2 - два значения переменной x, такие, что величины $q_1 = p^*(k=1\ /\ x_1) - p^*(k=1)$ и $q_2 = p^*(k=1\ /\ x_2) - p^*(k=1)$ имеют различные знаки в том смысле, что $q_1 \cdot q_2 \leq 0$. В таком случае величины $r_1 = p(k=1\ /\ x_2) - p(k=1)$ и $r_2 = p(k=1\ /\ x_2) - p(k=1)$ также имеют различные знаки в том смысле, что $r_1 \cdot r_2 \leq 0$.

Доказательство. Поскольку величины q_1 и q_2 имеют разные знаки, то вероятность $p^*(k=1)$ есть выпуклая линейная комбинация вероятностей $p^*(k=1\ /\ x_1.)$ и $p^*(k=1\ /\ x_2.)$, то есть существует такое число β, $0 \leq \beta \leq 1$, что для всех y

$$p^*(y) = \beta\cdot\ p^*(y\ /\ x_1.) + (1\text{-}\beta)\cdot\ p^*(y\ /\ x_2.)\ . \tag{48}$$

Из того факта, что равенство (40) выполняется для всех x и всех k, следует, что равенство

$$\sum_y \frac{p^*(y\ /\ x)}{p(y\ /\ x)}\cdot\left[p(y\ /\ k=1) \text{ - } p(y\ /\ k=2)\right] = 0 \tag{49}$$

выполняется для всех x, а следовательно, в том числе для x_1 и x_2, т.е.

$$\sum_y \frac{p^*(y\ /\ x_1)}{p(y\ /\ x_1)} \cdot \left[p(y\ /\ k=1) - p(y\ /\ k=2)\right] = 0 \quad , \tag{50}$$

$$\sum_y \frac{p^*(y\ /\ x_2)}{p(y\ /\ x_2)} \cdot \left[p(y\ /\ k=1) - p(y\ /\ k=2)\right] = 0. \tag{51}$$

Очевидно также, что

$$\sum_y \frac{p^*(y)}{p(y)} \cdot \left[p(y\ /\ k=1) - p(y\ /\ k=2)\right] = 0 \quad , \tag{52}$$

т.к. ранее (см. п. 1) было уже доказано, что $p^*(y) = p(y)$.

Допустим для определенности, что $p(k=1\ /\ x_1.) < p(k=1\ /\ x_2.)$ и докажем, что в этом случае вероятность $p(k=1)$ лежит в интервале от $p(k=1\ /\ x_1)$ до $p(k=1\ /\ x_2)$, то есть, что величины $p(k=1\ /\ x_1) - p(k=1)$ и $p(k=1\ /\ x_2) - p(k=1)$ имеют разные знаки.

Рассмотрим два выражения, зависящие от некоторого переменного коэфициента γ :

$$\sum_y \frac{p^*(y\ /\ x_1) \cdot \left[p(y\ /\ k=1) - p(y\ /\ k=2)\right]}{p(y\ /\ k=2) + \gamma \cdot \left[p(y\ /\ k=1) - p(y\ /\ k=2)\right]} \tag{53}$$

и

$$\sum_y \frac{p^*(y\ /\ x_2) \cdot \left[p(y\ /\ k=1) - p(y\ /\ k=2)\right]}{p(y\ /\ k=2) + \gamma \cdot \left[p(y\ /\ k=1) - p(y\ /\ k=2)\right]} \quad . \tag{54}$$

Как величина (53), так и величина (54), обе они строго убывают при увеличении коэффициента γ , т.к. по крайней мере для некоторых значений y $p(y / k=1) \neq p(y / k=2)$.

Величина (53) при $\gamma = p(k=1\ /\ x_1)$ равна (50), а при $\gamma = p(k=1\ /\ x_2)$ равна (51). Следовательно, если учесть, что величина (53) строго монотонно зависит от γ , то при всех $\gamma < p(k=1/ x_1)$ справедливы неравенства

$$\sum_y \frac{p^*(y/ x_1) \cdot \left[p(y/ k=1) - p(y/ k=2)\right]}{p(y/ k=2) + \gamma \cdot \left[p(y/ k=1) - p(y/ k=2)\right]} > 0 \quad , \tag{55}$$

$$\sum_y \frac{p^*(y/ x_2) \cdot \left[p(y/ k=1) - p(y/ k=2)\right]}{p(y/ k=2) + \gamma \cdot \left[p(y/ k=1) - p(y/ k=2)\right]} > 0 \quad , \tag{56}$$

а при всех $\gamma > p(k=2/ x_1)$ выполняются неравенства

$$\sum_y \frac{p^*(y/x_1)\cdot\left[p(y/k=1)-p(y/k=2)\right]}{p(y/k=2)+\gamma\cdot\left[p(y/k=1)-p(y/k=2)\right]} < 0 \ , \tag{57}$$

$$\sum_y \frac{p^*(y/x_2)\cdot\left[p(y/k=1)-p(y/k=2)\right]}{p(y/k=2)+\gamma\cdot\left[p(y/k=1)-p(y/k=2)\right]} < 0 \ . \tag{58}$$

Учитывая (48), складывая неравенства (55) и (56) с весами β и $1-\beta$, получаем, что для всех $\gamma < p(k=1\ /x_1)$

$$\sum_y \frac{p^*(y)\cdot\left[p(y/k=1)-p(y/k=2)\right]}{p(y/k=2)+\gamma\cdot\left[p(y/k=1)-p(y/k=2)\right]} > 0 \ , \tag{59}$$

а для всех $\gamma > p(k=1/x_2)$

$$\sum_y \frac{p^*(y)\cdot\left[p(y/k=1)-p(y/k=2)\right]}{p(y/k=2)+\gamma\cdot\left[p(y/k=1)-p(y/k=2)\right]} < 0 \ . \tag{60}$$

При $\gamma = p(k=1)$ выражения в левых частях (59) и (60) становятся равными левой части (52) и в силу (52) становятся равными нулю. Следовательно $p(k=1)$ обязательно лежит в интервале от $p(k=1/x_1)$ до $p(k=1/x_2)$, т.к. в любом другом случае левые части (59) и (60) не равны нулю. Таким образом разности $p(k=1/x_1)-p(k=1)$ и $p(k=1/x_2)-p(k=1)$ имеют разные знаки. **Лемма доказана.**

5. Лемма 2. Для любого x разность $p(k=1/x)-p(k=1)$ имеет тот же знак, что и разность $p(x/k=1)-p(x/k=2)$.

Доказательство. Пусть $p(x/k=1)=p(x/k=2)$. В таком случае

$$p(k=1/x)=\frac{p(k=1)\cdot p(x/k=1)}{p(k=1)\cdot p(x/k=1)+p(k=2)\cdot p(x/k=2)} = \frac{p(k=1)}{p(k=1)+p(k=2)} = p(k=1) \ .$$

Пусть $p(x/k=1)>p(x/k=2)$. Введя обозначение $\gamma = \dfrac{p(x/k=1)}{p(x/k=2)} > 1$, получаем

$$p(k=1/x)=p(k=1)\cdot\frac{1}{p(k=1)+\dfrac{p(k=2)}{\gamma}} > p(k=1) \ .$$ **Лемма доказана.**

6. Из доказанных лемм вытекает следствие.

Следствие. Пусть x^* некоторое значение переменной x, такое, что $p^*\left(k=1/x^*\right)\neq p^*(k=1)$. В таком случае величина

$$\left[\,p(x/k=1)-p(x/k=2)\right]\frac{p^{*}(k=1/x)-p^{*}(k=1)}{p^{*}\left(k=1/x^{*}\right)-p^{*}(k=1)} \tag{61}$$

либо для всех $x\in X$ является неположительной, либо для всех $x\in X$ является неотрицательной.

Доказательство. Действительно, знаменатель в выражении (61) не зависит от x. Следовательно, если при изменении x величина (61) меняет свой знак, то меняет свой знак одна и только одна из двух величин: либо $p(x/k=1)-p(x/k=2)$, либо $p(k=1/x)-p(k=1)$, а это противоречит либо лемме 1, либо лемме 2. **Следствие доказано.**

7. Поскольку равенство (39) справедливо как для $k=1$, так и для $k=2$, то из него вытекают равенства

$$\sum_{x}\frac{p^{*}(x/y)}{p(x/y)}\left[p(x/k=1)-p(x/k=2)\right]=0\ , \tag{62}$$

$$\sum_{x}\frac{p^{*}(x/y)}{p(x/y)}p(x)=1\ . \tag{63}$$

8. Пусть x^{*} - такое значение x, что по крайней мере для некоторых значений y выполняется неравенство $p^{*}\left(y/x^{*}\right)\neq p^{*}(y)$. Далее доказательство должно быть выполнено отдельно для двух случаев: когда такое x^{*} существует и когда оно не существует. Допустим, что такое x^{*} существует, и зафиксируем его для дальнейшего рассмотрения. Случай, когда такого значения не существует, будет рассмотрен далее, в п. 11.

Поскольку (62) справедливо для любого значения y, то справедливо и равенство

$$\sum_{y}\left[p^{*}\left(y/x^{*}\right)-p^{*}(y)\right]\sum_{x}\frac{p^{*}(x/y)}{p(x/y)}\left[p(x/k=1)-p(x/k=2)\right]=0\ ,$$

которое легко преобразуется к виду

$$\sum_{x}\left[p(x/k=1)-p(x/k=2)\right]\sum_{y}\frac{p^{*}(y/x)}{p(y/x)}\left[p^{*}\left(y/x^{*}\right)-p^{*}(y)\right]=0\ . \tag{64}$$

При этом преобразовании принималось во внимание, что справедлива цепочка равенств

$$\frac{p^{*}(x/y)}{p(x/y)}=\frac{p^{*}(y)\cdot p^{*}(y/x)}{p(y)\cdot p(y/x)}=\frac{p^{*}(x)\cdot p^{*}(y/x)}{p(x)\cdot p(y/x)}=\frac{p^{*}(y/x)}{p(y/x)}\ ,$$

которая верна в силу того, что $p^{*}(x)=p(x)$ и $p^{*}(y)=p(y)$.

9. Исследуем сумму $f(x)=\sum_{y}\frac{p^{*}(y/x)}{p(y/x)}\left[p^{*}\left(y/x^{*}\right)-p^{*}(y)\right]$. Разобьем множество X на два подмножества: $X^{=}$ и $X^{\neq}$. Первое из этих множеств включает в себя те

значения x, для которых $p^*(y/x) = p^*(y)$, а следовательно, $p(y/x) = p(y)$. Второе из этих множеств, т.е. $X^{\neq}$, содержит все прочие значения x.

Для каждого $x \in X^{=}$ справедливо

$$f(x) = \sum_y \frac{p^*(y)}{p(y)}\left[p^*\left(y/x^*\right) - p^*(y)\right] = \sum_y p^*\left(y/x^*\right) - \sum_y p^*(y) = 0 \ .$$

Для всех прочих значений x справедливо в силу (47)

$$f(x) = \frac{p^*\left(k=1/x^*\right) - p^*(k=1)}{p^*(k=1/x) - p^*(k)} \sum_y \frac{p^*(y/x)}{p(y/x)}\left[p^*(y/x) - p^*(y)\right] \ .$$

Подставив это выражение в (64), получаем

$$\sum_{x \in X^{\neq}}\left[p(x/k=1) - p(x/k=2)\right]\frac{p^*\left(k=1/x^*\right) - p^*(k=1)}{p^*(k=1/x) - p^*(k=1)} \sum_y \frac{p^*(y/x)}{p(y/x)}\left[p^*(y/x) - p^*(y)\right] = 0 \quad (65)$$

10. Исследуем свойства суммы $g(x) = \sum_y \frac{p^*(y/x)}{p(y/x)}\left[p^*(y/x) - p^*(y)\right]$, входящей в выражение (65).

Лемма 3. Справедливо неравенство

$$g(x) = \sum_y \frac{p^*(y/x)}{p(y/x)}\left[p^*(y/x) - p^*(y)\right] \geq 0 \ ,$$

причем равенство соблюдается тогда и только тогда, когда для любого y справедливо равенство $p^*(y/x) = p(y/x)$.

Доказательство. Учитывая, что $p^*(y) = p(y)$

$$g(x) = \sum_y \frac{p^*(y/x)}{p(y/x)} p^*(y/x) - \sum_y \frac{p^*(y/x)}{p(y/x)} p(y) \ .$$

Ранее уже было доказано, что

$$\sum_y \frac{p^*(y/x)}{p(y/x)} p(y) = 1 \, . \quad (66)$$

Докажем, что

$$\sum_y \frac{p^*(y/x)}{p(y/x)} p^*(y/x) \geq 1 \ , \quad (67)$$

и выясним условия, при которых (67) обращается в равенство. Для этого найдем минимум левой части (67) по значениям $p(y/x)$ при условии $\sum_y p(y/x)=1$. Функция Лагранжа, соответствующая этой оптимизационной задаче, есть $H=\sum_y \frac{p^*(y/x)}{p(y/x)}p^*(y/x)+\lambda\sum_y p(y/x)$, а искомый минимум есть решение системы уравнений

$$\frac{\partial H}{\partial p(y/x)}=-\frac{p^*(y/x)\cdot p^*(y/x)}{p^2(y/x)}+\lambda=0,\ y\in Y\ ,$$

которая имеет единственное решение $p(y/x)=p^*(y/x)$. Подставляя это решение в правую часть (67), получаем, что минимальное значение этой правой части есть 1, что доказывает (67), а, следовательно, и лемму в целом. **Лемма доказана.**

11. Исследуем равенство (65), записав его в краткой форме

$$\sum_{x\in X^{\neq}} h(x)\cdot f(x)=0\,, \tag{68}$$

где

$$h(x)=\left[p(x/k=1)-p(x/k=2)\right]\frac{p^*\left(k=1/x^*\right)-p^*(k=1)}{p^*(k=1/x)-p^*(k=1)}\ ,$$

$$f(x)=\sum_y \frac{p^*(y/x)}{p(y/x)}\left[p^*(y/x)-p^*(y)\right]\ .$$

В сумме (68) в силу следствия из лемм 1 и 2 (см.п.6) все величины $h(x)$ имеют какой-то один знак: либо все они неположительны, либо все они неотрицательны. В силу только что доказанной леммы все величины $f(x)$ неотрицательны. Следовательно, из равенства суммы $\sum_{x\in X^{\neq}} h(x)\cdot f(x)$ нулю следует равенство нулю и всех ее слагаемых. Это значит, что $\sum_y \frac{p^*(y/x)}{p(y/x)}\left[p^*(y/x)-p^*(y)\right]=0$ для всех $x\in X^{\neq}$, что возможно лишь тогда, когда $p(y/x)=p^*(y/x)$ для всех $y\in Y, x\in X^{\neq}$. Для значений $x\in X^{\neq}$ справедливо по определению, что $p^*(y/x)=p^*(y)$ и $p(y/x)=p(y)$. Следовательно, равенство $p^*(y/x)=p(y/x)$ выполняется и для $x\in X^{\neq}$.

11. Мы доказали, что равенство

$$p^*(y/x)=p(y/x) \tag{69}$$

выполняется для всех $y\in Y$ и $x\in X$ в случае, когда существует такое значение x^*, что по крайней мере для некоторых значений y выполняется неравенство $p^*\left(y/x^*\right)\neq p^*(x)$.

Рассмотрим теперь случай, когда такого значения x^* не существует. Это значит, что для всех x и y выполняется равенство $p^*(y/x) = p^*(y)$, отсюда следует, что для всех x и y выполняется и равенство $p^*(x/y) = p^*(x)$. Условия теоремы в этом случае приобретают вид

$$\sum_x \frac{p^*(x)}{p(x/y)} p(x/k) = 1, \quad \sum_y \frac{p^*(y)}{p(y/x)} p(y/k) = 1 \; .$$

Поскольку эти равенства выполняются как для $k=1$, так и для $k=2$, то из них следуют равенства $\sum_x \frac{p^*(x)}{p(x/y)} p(x) = 1$ и $\sum_y \frac{p^*(y)}{p(y/x)} p(y) = 1$. Учитывая, что $p(x) = p^*(x)$, а $p(y) = p^*(y)$, получаем равенства $\sum_y \frac{p^*(y) \cdot p^*(y)}{p(y/x)} = 1$ и $\sum_x \frac{p^*(x) \cdot p^*(x)}{p(x/y)} = 1$, которые в силу доказанной леммы 3 могут выполняться лишь в том случае, когда для всех x и y выполняется равенство $p(y/x) = p^*(y)$ и $p(x/y) = p^*(x)$.

12. Таким образом, мы доказали, что из условий теоремы следует, что для всех x выполняется $p(x) = p^*(x)$ и для всех x и y выполняется $p(y/x) = p^*(y/x)$. Следовательно, для всех x и y выполняется $\sum_k p(k) \cdot p(x/k) \cdot p(y/x) = p^*(x,y)$, из чего достаточно очевидно следует неравенство (36), которое требовалось доказать. **Теорема доказана.**

Доказанная теорема существенно углубляет наши знания о поведении алгоритмов идентификации смеси по сравнению с имеющимися до настоящего времени представлениями. Теоретические результаты, которые имелись до настоящего времени, сводятся, в основном, к тому, что глобальный максимум функции правдоподобия есть одна из неподвижных точек алгоритма. Эти результаты не выходят за рамки результатов, полученных уже в самой первоначальной публикации [1] об этих алгоритмах. Очевидно, что имеются еще и другие неподвижные точки, которые явно не являются глобальными максимумами, то есть соответствуют ложным решениям алгоритма. Такими являются, например, решения, когда $p(k) = 0$ для некоторого k или когда $p(x \; / \, k = 1) = p(x \; / \, k = 2)$, или тому подобные ситуации. Отсутствие до настоящего времени исчерпывающего описания множества всех возможных неподвижных точек оставляло место для существенного неудовлетворения в чисто научном плане и дискомфорта при практических применениях. Понимание одного лишь того факта, что неподвижная точка есть локальный максимум, никоим образом этот дискомфорт не снимает, так как в каждом конкретном случае применения пользователь остается в неведении, существует ли решение, лучшее, чем найденное алгоритмом, или нет.

В настоящее же время, в результате выполненного исследования достигнута следующая ясность. Неподвижной точкой алгоритма является такое и только такое решение $p(k),\ p(x\ /\ k), p(y\ /\ k)$, для которого

а). либо $p(k) = 0$ для некоторого k ;

б). либо $p(x\ /\ k) = p(x)$ и $p(y\ /\ k) = p(y)$ для всех $x,\ y,\ k$;

в). либо решение $p(k),\ p(x\ /\ k), p(y\ /\ k)$ обеспечивает глобальный максимум функционала $\sum\limits_{x,y} p^*(x,y) \cdot \log \sum\limits_{k} p(k) \cdot\ p(x\ /\ k) \cdot p(y\ /\ k)$.

Следовательно, если для решения, выданное алгоритмом, не выполняются легко распознаваемые условия а) и б), то это решение является глобальным максимумом функции правдоподобия. Этот вывод является основным положительным результатом выполненного исследования.

Литература

1. Шлезингер М.И. Взаимосвязь обучения и самообучения в распознавании образов // Кибернетика.- 1968.- № 2.- С. 81-88.

2. Шлезингер М.И. Математические средства обработки изображений.- Киев: Наук. думка, 1989. - 198 с.