ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

УДК 519.233.33+519.244.4

СТАТИСТИЧЕСКИЕ МЕТОДЫ ПОИСКА НАБОРА КООРДИНАТ, НА КОТОРОМ СЛУЧАЙНЫЙ ВЕКТОР ИМЕЕТ ЗАПРЕТЫ

О. В. Денисов

ООО «Центр сертификационных исследований», г. Москва, Россия

Наблюдается стационарная последовательность случайных векторов длины L, имеющих распределение случайного вектора ξ; координаты векторов принимают значения в конечном множестве. Рассматривается гипотеза о существовании некоторого множества номеров координат $\Theta \subset \{1, \ldots, L\}$, такого, что подвектор ξ_Θ (проекция ξ на координаты с номерами из Θ) распределён как заданный случайный вектор η, распределение которого имеет запреты. Строится критерий согласия на основе анализа запретов эмпирического распределения. Когда априори известно, что гипотеза выполнена, предлагаются три алгоритма поиска части Θ, работающие при разной доле информации о распределении случайного вектора η.

Ключевые слова: *статистический критерий, запреты распределений.*

DOI 10.17223/20710410/28/1

STATISTICAL METHODS OF SEARCH FOR COORDINATE SET ON WHICH A RANDOM VECTOR HAS BANS

O. V. Denisov

Certification Research Center, Moscow, Russia

E-mail: denisovOleg@yandex.ru

A stationary sequence of random vectors of length L with the distribution of a random vector ξ is observed. Coordinates of vectors in it take values in a finite set. The following hypothesis is considered: there is a set $\Theta \subset \{1, \ldots, L\}$ such that the subvector ξ_Θ (being the projection of ξ onto coordinates with numbers in Θ) has the distribution of a given random vector η with the distribution having bans. A concordance criterion is constructed by the analysis of an empirical distribution bans. In the case of the hypothesis validity (a priori), three algorithms to search for a part of Θ are proposed. They work under various portions of the information about the random vector η distribution.

Keywords: *statistical test, bans of distributions.*

Введение

Понятие запретов дискретного вероятностного распределения введено в рабо тах [1, 2]; в [2] предложен статистический метод поиска минимальных запретов. В частности, такие методы могут иметь приложения в стеганографии [1, пример 3].

В данной работе эти подходы применяются при решении более конкретно поставленной задачи проверки статистической гипотезы о наличии некоторого частного распределения заданного вида у наблюдаемого многомерного распределения. Для этого развивается понятийный аппарат, связанный с запретами. Главными новыми понятиями являются «графы простых запретов» и «устойчивость координаты случайного вектора к запретам». На их основе построен критерий согласия с гипотезой и три алгоритма поиска некоторых номеров координат из тех, на которых сосредоточено частное распределение; получены оценки вероятностей ошибок критерия и алгоритмов.

Заметим, что ранее запреты распределения выхода простейших неавтономных автоматов исследовались в связи с криптографическими приложениями [3, 4]. При этом определение запрета давалось несколько иначе — в терминах решений систем автоматных уравнений. Для двоичного регистра сдвига понятие запрета введено С. Н. Сумароковым в 1968 г.

Перейдём к строгой постановке задачи. Пусть X — произвольное конечное множество мощности k. Наблюдается отрезок стационарной последовательности случайных векторов длины L

$$\mathbf{x}(t) = (x_1(t), \dots, x_L(t)) \in X^L, \quad 1 \leqslant t \leqslant N,$$

имеющих распределение случайного вектора ξ: $\mathbf{x}(t) \sim \xi$, $1 \leqslant t \leqslant N$, N — длина отрезка (число наблюдений).

Рассматривается задача проверки сложной статистической гипотезы $H(\eta)$ о наличии *особенного* множества номеров координат

$$\Theta = \{\theta_1, \dots, \theta_M\} \subset \{1, \dots, L\},$$

такого, что соответствующий подвектор на этих координатах ξ_Θ имеет распределение случайного вектора

$$\eta = (\eta_1, \dots, \eta_M),$$

известной длины $M < L$, которое полностью или частично известно. Далее предполагается, что оно имеет запреты; строгое определение запрета распределения приведено ниже.

Через $\mathbf{x}_I = (x_{i_1}, \dots, x_{i_r})$ здесь и далее обозначаем подвектор вектора $\mathbf{x}$, состоящий из координат с номерами из множества $I = \{i_1, \dots, i_r\}$. Для произвольного множества A обозначим через 2^A множество всех подмножеств множества A; $A^{(r)}$ — множество всех r-элементных подмножеств A; $\bar{A}$ — дополнение к множеству A.

Итак, гипотеза $H(\eta)$ формулируется как гипотеза о том, что ξ имеет соответствующее частное распределение:

$$H(\eta) = \Big\{ \exists\, \Theta \subset \{1, \dots, L\}^{(M)} : \ \xi_{\boldsymbol{\Theta}} \sim \eta \Big\}.$$

Если априори известно, что предположение $H(\eta)$ выполнено, то встаёт вопрос об оценке множества Θ как параметра распределения.

Далее в п. 1 введён ряд новых понятий, связанных с носителями и запретами частных распределений, попутно доказывая некоторые их свойства. В п. 2 построен статистический критерий проверки гипотезы $H(\eta)$ и при некоторых предположениях получены оценки его ошибок. Затем предлагаются три алгоритма поиска части множества Θ, работающие при разной доле информации о распределении η. При полной информации о η предлагается последовательный алгоритм 1; он имеет нулевую вероятность ошибки. Алгоритмы 2 и 3 строятся при отсутствии информации о распределении η и работают на фиксированном объёме материала.

В п. 3 вводятся ограничения на распределение ξ, упрощающие построение критерия, расчёт параметров критерия и алгоритмов. Для схемы независимых наблюдений получены верхние оценки для числа наблюдений N, при котором вероятности ошибок не превосходят заданной величины.

Согласно [2, с.57], данные методы могут быть применены для «...статистического выявления скрытых каналов, в которых вставки осуществляются с помощью некоторых функциональных соотношений». У нас это соответствует задаче поиска распределения вида «вектор аргументов и вектор-функция от него» при фиксированной $\mathbf{f}(\mathbf{x}) : X^n \to X^m$, где вектор аргументов $\mathbf{x}$ не имеет запретов. В простейшем случае, когда $\mathbf{x} = (x_1, \ldots, x_n)$ распределён равномерно, случайный вектор обозначаем через

$$\mathbf{xf} = (x_1, \ldots, x_n, f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})), \quad \mathbf{x} \sim U(X^n).$$

1. Основные понятия

Введём сначала теоретические характеристики распределения ξ, связанные с запретами, а затем эмпирические.

1.1. Запреты распределения случайного вектора и графы запретов

Случайный вектор ξ имеет дискретное распределение, поэтому далее без ограничения общности считаем, что *носитель распределения* ξ

$$\mathrm{Supp}(\xi) = \left\{\mathbf{x} \in X^L : \mathbf{P}\{\xi = \mathbf{x}\} > 0\right\}$$

совпадает с множеством значений случайного вектора: $\mathrm{Supp}(\xi) = \xi(\Omega)$, то есть ξ принимает все свои значения с положительной вероятностью. Разрабатываемые методы основаны на анализе только наблюдаемых носителей частных распределений ξ.

Введём ряд определений и обозначений. Вектор $\mathbf{a} \in X^L$ называется *запретом размерности L распределения* ξ, если $\mathbf{P}\{\xi = \mathbf{a}\} = 0$. При этом $\mathbf{a}$ называется *простым запретом размерности* L, если

$$\mathbf{P}\left\{\xi_{\{1,\ldots,L\}\setminus\{i\}} = \mathbf{a}_{\{1,\ldots,L\}\setminus\{i\}}\right\} > 0,\ 1 \leqslant i \leqslant L,$$

то есть любой его собственный подвектор не является запретом распределения любого подвектора ξ.

Множество всех запретов размерности L распределения ξ обозначим через $\mathcal{Z}(\xi)$, а простых запретов размерности L — через $\mathcal{Z}_s(\xi)$. Очевидно, что

$$\mathcal{Z}_s(\xi) \subset \mathcal{Z}(\xi) = X^L \setminus \mathrm{Supp}(\xi).$$

Принадлежность вектора к запретам может определяться лишь значениями некоторых его координат. Чтобы далее выявлять такие наборы координат, рассмотрим запреты частных распределений.

Вектор $\mathbf{a} \in X^r$ будем называть *запретом размерности r распределения ξ на множестве (номеров координат)* $J \in \{1, \dots, L\}^{(r)}$, если $\mathbf{a} \in \mathcal{Z}(\xi_J)$, $1 \leqslant r \leqslant L$. При этом $\mathbf{a}$ будем называть *простым запретом распределения ξ (размерности r)*, если $\mathbf{a} \in \mathcal{Z}_s(\xi_J)$. Это равносильно тому, что для $J = \{j_1, \dots, j_r\}$ выполнено

$$\mathbf{P}\{\xi_J = \mathbf{a}\} = 0, \qquad \mathbf{P}\left\{\xi_{J\setminus\{j_s\}} = \mathbf{a}_{\{1,\dots,r\}\setminus\{s\}}\right\} > 0,\ 1 \leqslant s \leqslant r.$$

Кратчайшими запретами распределения ξ будем называть запреты наименьшей размерности. Эту размерность назовём *устойчивостью к запретам* распределения ξ и обозначим $\mathbf{z}_{\min}(\xi)$.

Очевидно, что кратчайшие запреты всегда являются простыми. Из определений также следует, что свойство «быть простым запретом» сохраняется при расширении случайного вектора: для $|J| = r$, $\mathbf{a} \in X^r$ выполнено

$$\begin{gathered}\mathbf{a} \in \mathcal{Z}_s(\xi_J) \text{ тогда и только тогда, когда } \mathbf{a} \text{ является простым запретом} \\ \text{размерности } r \text{ распределения } \xi_I \text{ для всех } I \supset J.\end{gathered} \tag{1}$$

Заметим, что каждый запрет размерности r на наборе J влечёт появление k запретов размерности $r+1$ на наборах вида $J \cup \{i\}$, $i \notin J$. Такие запреты размерности $r+1$ не несут новой информации о множестве $\mathrm{Supp}(\xi)$ при знании всех запретов размерности r, и в этом смысле интересны лишь простые запреты.

Теперь введём основные инструменты анализа.

Определение 1. *Графом запретов размерности r распределения ξ* назовём r-однородный гиперграф (кратко r-граф) на вершинах с номерами из $\{1, \dots, L\}$

$$\mathcal{G}(r, \xi) = \left\{J \subset \{1, \dots, L\}^{(r)} : \mathcal{Z}_s(\xi_J) \neq \varnothing\right\}.$$

Его рёбрами являются все наборы, на которых распределение ξ имеет простые запреты размерности r. Граф

$$\mathcal{G}_{\min}(\xi) = \mathcal{G}(\mathbf{z}_{\min}(\xi), \xi)$$

назовём *графом кратчайших запретов*. Его можно определить также как первый непустой r-граф в цепочке $\mathcal{G}(1, \xi)$, $\mathcal{G}(2, \xi)$, …, $\mathcal{G}(L, \xi)$.

Везде далее будем считать, что $\mathcal{Z}(\xi) \neq \varnothing$ (в противном случае анализ не даст никакой информации о распределении ξ), и тогда в этой цепочке есть хотя бы один непустой r-граф.

Обозначим также через

$$\mathcal{G}(\xi) = \bigsqcup_{1 \leqslant r \leqslant L} \mathcal{G}(r, \xi) \subset 2^{\{1,\dots,L\}}$$

гиперграф (неоднородный в общем случае), состоящий из наборов всех простых запретов.

Для произвольного гиперграфа $\mathcal{G} \subset 2^{\{1,\dots,L\}}$ через

$$\mathbf{V}(\mathcal{G}) = \bigcup_{J \in \mathcal{G}} J, \quad \mathbf{v}(\mathcal{G}) := |\mathbf{V}(\mathcal{G})|$$

обозначим соответственно множество тех вершин гиперграфа $\mathcal{G}$, которые покрыты хотя бы одним его ребром, и число таких вершин.

Определение 2. Будем говорить, что r-графы $\mathcal{G}$ и $\mathcal{G}'$ *изоморфны* ($\mathcal{G} \cong \mathcal{G}'$), если существует биекция $\phi : \mathbf{V}(\mathcal{G}) \to \mathbf{V}(\mathcal{G}')$, такая, что

$$J \in \mathcal{G} \Longleftrightarrow \phi(J) \in \mathcal{G}.$$

Это означает, что $\mathcal{G}'$ может быть получен из $\mathcal{G}$ взаимно однозначным изменением номеров вершин после удаления из $\mathcal{G}$ и $\mathcal{G}'$ вершин, не лежащих ни в одном ребре. Необходимым условием изоморфности, очевидно, является равенство $\mathbf{v}(\mathcal{G}) = \mathbf{v}(\mathcal{G}')$.

Для r-графа $\mathcal{G}$ и множества $I \subset \{1, \ldots, L\}$, $1 \leqslant r \leqslant |I|$, через

$$\mathcal{G}_I = \mathcal{G} \cap I^{(r)}$$

обозначим ограничение $\mathcal{G}$ на множество вершин I, то есть соответствующий r-подграф $\mathcal{G}$. Из условия (1) имеем равенство $\mathcal{G}(r, \xi)_I = \mathcal{G}(r, \xi_I)$. Отсюда следует, что при гипотезе $H(\eta)$ имеем

$$\mathcal{G}(r, \xi)_\Theta = \mathcal{G}(r, \xi_\Theta) \cong \mathcal{G}(r, \eta), \quad 1 \leqslant r \leqslant M.$$

На этом свойстве основан критерий согласия с гипотезой $H(\eta)$ и статистические алгоритмы поиска частей множества Θ, в том числе его части:

$$\Theta_{\min} = \mathbf{V}(\mathcal{G}_{\min}(\xi_\Theta)) \subset \Theta.$$

Так как $\mathcal{G}_{\min}(\xi_\Theta) \cong \mathcal{G}_{\min}(\eta)$, то $|\Theta_{\min}| = \mathbf{v}(\mathcal{G}_{\min}(\eta))$.

1.2. Эмпирические характеристики распределения

Заметим, что введённые выше характеристики инвариантны относительно любого изменения распределения вероятностей ξ, при котором сохраняется носитель распределения. Поэтому их определения можно дать без использования понятия вероятности, оперируя лишь множеством $\mathrm{Supp}(\xi)$.

Продемонстрируем этот путь при определении аналогичных эмпирических характеристик распределения ξ. Обозначим через

$$\mathcal{X} = \mathcal{X}(N) \subset \mathrm{Supp}(\xi) \subset X^L$$

множество всех различных векторов среди наблюдений $\mathbf{x}(1), \ldots, \mathbf{x}(N)$. Элементы множества $\mathcal{Z}(\mathcal{X}) = X^r \backslash \mathcal{X}$ назовём *эмпирическими запретами* (*запретами множества* $\mathcal{X}$) *размерности* L.

Введём операцию ограничения $\mathcal{X}$ на множество координат с номерами из $I \in \{1, \ldots, L\}^{(r)}$ (*проекции* $\mathcal{X}$ на I): $\mathcal{X}_I = \{\mathbf{a}_I : \mathbf{a} \in \mathcal{X}\} \subset X^r$.

Вектор $\mathbf{a} \in \mathcal{Z}(\mathcal{X})$ называется *простым эмпирическим запретом размерности* L, если $\mathbf{a}_I \in \mathcal{X}_I$ для всех $\varnothing \neq I \subset \{1, \ldots, L\}$. Множество всех простых эмпирических запретов размерности L обозначим через $\mathcal{Z}_s(\mathcal{X})$.

Аналогично запретам распределения, для $J \in \{1, \ldots, L\}^{(r)}$ элементы множеств $\mathcal{Z}(\mathcal{X}_J)$ и $\mathcal{Z}_s(\mathcal{X}_J)$ называем *r-мерными эмпирическими запретами* (соответственно *простыми r-мерными эмпирическими запретами*) на множестве J; определяем *r-графы эмпирических запретов* $\mathcal{G}(r, \mathcal{X})$, $1 \leqslant r \leqslant L$, и их объединение — гиперграф $\mathcal{G}(\mathcal{X})$.

Заметим, что все эти понятия можно было ввести сразу на базе предыдущих определений, формально интерпретируя параметр $\mathcal{X}$ как обозначение для некоторого произвольного распределения с носителем $\mathcal{X}$.

Множества $\mathcal{X}$ и $\mathcal{X}_I$ являются статистическими оценками носителей $\mathrm{Supp}(\xi)$ и $\mathrm{Supp}(\xi_{\mathbf{I}})$, всегда лежащими в них, и поэтому всегда $\mathcal{Z}(\mathcal{X}_I) \supset \mathcal{Z}(\xi_I)$, $I \subset \{1,\dots,L\}$.

Легко видеть, что

$$\begin{gathered}\text{всегда } \mathbf{z}_{\min}(\mathcal{X}) \leqslant \mathbf{z}_{\min}(\xi), \text{ и если здесь достигается равенство,} \\ \text{то } \mathcal{G}_{\min}(\mathcal{X}) \supset \mathcal{G}_{\min}(\xi).\end{gathered} \tag{2}$$

Для множеств простых запретов включение $\mathcal{G}(r,\mathcal{X}(N)) \supset \mathcal{G}(r,\xi)$ в общем случае неверно, что показывает следующий

Пример 1. Пусть $X = \mathbb{Z}_2$, $L = 2$, $\mathcal{Z}(\xi) = \{(1,0)\}$. Тогда одномерных запретов ξ не имеет, $\mathcal{G}(1,\xi) = \varnothing$, а имеющийся запрет размерности 2 является простым, и граф $\mathcal{G}_{\min}(\xi) = \mathcal{G}(2,\xi)$ состоит из единственного ребра $\{1,2\}$.

Далее, пусть $\mathcal{X} = \{(0,0)\}$. Тогда эмпирическими запретами размерности 2 являются все ненулевые векторы, вектор (1) является простым запретом на множествах $\{1\}$ и $\{2\}$, $\mathcal{G}(1,\mathcal{X}) = \{\{1\},\{2\}\} \supset \mathcal{G}(1,\xi)$. Отсюда также следует, что любой ненулевой вектор длины 2 не является простым эмпирическим запретом, и поэтому $\mathcal{Z}_s(\mathcal{X}) = \varnothing$, $\mathcal{G}(2,\mathcal{X}) = \varnothing$.

Этот пример также показывает существенность условия о равенстве в (2).

1.3. Предположение о вероятностной модели

Далее будем считать, что для любых $0 < \alpha < 1$ и $1 \leqslant R \leqslant L$ определена функция $N_1 = N_1(R,\xi,\alpha)$, такая, что

$$\mathbf{P}\left\{\forall J \in \{1,\dots,L\}^{(R)}\ \ (\mathcal{X}(N)_J = \mathrm{Supp}(\xi_J))\right\} \geqslant 1-\alpha \tag{3}$$

при $N \geqslant N_1(R,\xi,\alpha)$. Это означает, что при всех достаточно больших N с вероятностью не менее заданной все R-мерные проекции выборки принимают все возможные для них значения. Заметим, что функция N_1 зависит не только от распределения ξ в фиксированный момент времени, но и от распределения $\{\mathbf{x}(t)\}_{t\geqslant 1}$ всей последовательности состояний. Для упрощения обозначений эту зависимость будем иметь в виду, не указывая явно.

Очевидно, что при $N \geqslant N_1(R,\xi,\alpha)$ справедливы следующие оценки:

$$\begin{gathered}\mathbf{P}\left\{\forall I, 1 \leqslant |I| \leqslant R\ \ (\mathcal{Z}(\mathcal{X}(N)_I) = \mathcal{Z}(\xi_I), \mathcal{Z}_s(\mathcal{X}(N)_I) = \mathcal{Z}_s(\xi_I))\right\} \geqslant 1-\alpha; \\ \mathbf{P}\left\{\forall\, r \in \{1,\dots,R\}\ (\mathcal{G}(r,\mathcal{X}(N)) = \mathcal{G}(r,\xi))\right\} \geqslant 1-\alpha.\end{gathered} \tag{4}$$

1.4. Гипотеза о наличии функциональной вставки

Важным частным случаем гипотезы $H(\eta)$ является случай, когда η имеет распределение вида «вектор аргументов и функции от него», где вектор аргументов не имеет запретов. В частности, это выполнено для $\eta \sim \mathbf{xf}$. Тогда при фиксированной $\mathbf{f} : X^n \to X^m$ сложная статистическая гипотеза $H(\eta)$ строго формулируется так:

$$\exists \Theta = Inp \sqcup Out \in \{1,\dots,L\}^{(n+m)}\ \Big(\xi(\Omega)_{Inp} = X^n,\ \forall \omega \in \Omega\ \left(\xi(\omega)_{Out} = \mathbf{f}\left(\xi(\omega)_{Inp}\right)\right)\Big). \tag{5}$$

Здесь множества Inp, Out образуют разбиение множества Θ; первое может являться множеством номеров аргументов, а второе — множеством номеров значений функции.

Заметим, что гипотеза $H(\eta)$ является существенным обобщением (5) и включает также случаи, когда на координатах с номерами особенного множества реализована не вектор-функция, а нечто более сложное.

Рассмотрим пример, демонстрирующий идею использования графов запретов при поиске частных распределений вида $\mathbf{xf}$.

Пример 2. Пусть $X = \mathbb{Z}_2$, $L \geqslant 5$ и по наблюдениям над реализациями случайного вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_{L-2}, \xi_{L-1} = \xi_1\xi_2, \xi_L = \xi_1 \oplus \xi_2 \oplus \xi_3)$, $(\xi_1, \ldots, \xi_{L-2}) \sim U(\mathbb{Z}_2^{L-2})$, требуется найти функциональную вставку-конъюнкцию.

Решение. Здесь моделью частного распределения на особенном множестве координат является вероятностная схема $\eta = \mathbf{xf} = (x_1, x_2, x_1x_2)$, $(x_1, x_2) \sim U(\mathbb{Z}_2^2)$ вида $\mathbf{xf}$. Для неё $\mathcal{G}(1, \mathbf{xf}) = \varnothing$ и $\mathcal{G}_{\min}(\mathbf{xf}) = \mathcal{G}(2, \mathbf{xf}) = \big\{\{1, 3\}, \{2, 3\}\big\}$.

Одномерные распределения ξ также не имеют запретов, двумерные распределения имеют запрет (0,1) на наборах координат из $\mathcal{G}(2, \xi) = \big\{\{1, L-1\}, \{2, L-1\}\big\} \cong \mathcal{G}(2, \mathbf{xf})$. Тогда при $N \geqslant N_1(r, \xi, \alpha)$, $r = 2$ с вероятностью не менее $1-\alpha$, согласно (4), выполнено $\mathcal{G}(2, \mathcal{X}) = \mathcal{G}(2, \xi)$. Это позволяет сделать вывод, что вставка-конъюнкция возможна только на координатах с номерами из $\Theta^* = \{1, 2, L-1\}$, причём $x_{L-1} = f(x_1, x_2)$.

Устойчивость к запретам координат случайного вектора

Заметим, что если в примере 2 положить $\xi_L = \xi_1 \oplus \xi_2$, то ξ на $J = \{L-1, L\}$ будет иметь простой запрет (1,1), поскольку $\xi_1\xi_2 = 1 \Rightarrow \xi_1 = \xi_2 = 1 \Rightarrow \xi_1 \oplus \xi_2 = 0$. Тогда граф $\mathcal{G}(2, \mathcal{X}(N))$ при всех N будет содержать ребро $\{L-1, L\}$. Это нарушит изоморфизм $\mathcal{G}(2, \mathcal{X}(N))$ и $\mathcal{G}(2, \mathbf{xf})$ при всех N, и метод определения Θ в указанном виде будет неприменим.

Чтобы избежать таких сложностей, введём ограничение на распределение ξ: все кратчайшие запреты распределения ξ образованы координатами с номерами из Θ. Оно соответствует естественному предположению о том, что сначала осуществляется поиск подвектора, наиболее уязвимого для нашего метода. Далее всюду будем считать предположение выполненным. Оно записывается проще с помощью следующих определений.

Устойчивостью к запретам i-й координаты случайного вектора ξ назовём величину

$$\mathbf{z}(i, \xi) = \min\{|J| : J \ni i, \mathcal{Z}_s(\xi_J) \neq \varnothing\}.$$

Она равна наименьшему значению r, при котором $i \in \mathcal{G}(r, \xi)$, то есть номер i участвует в простом запрете размерности r. Если i не участвует ни в одном простом запрете, то по определению считаем $\mathbf{z}(i, \xi) = \infty$.

Очевидно, что

$$\mathbf{z}_{\min}(\xi) = \min_{1 \leqslant i \leqslant L} \mathbf{z}(i, \xi).$$

Легко видеть, что в примере 2 справедлива импликация $(\xi_3\xi_{L-1} = 1 \Rightarrow \xi_L = 1)$, откуда следует, что $\{3, L-1, L\}$ — ребро графа $\mathcal{G}(3, \xi)$. Поэтому с учётом далее доказываемой теоремы 5, п. 3, имеем

$$\mathbf{z}(i, \xi) = \begin{cases} 2 & \text{при } i \in \{1, 2, L-1\}, \\ 3 & \text{при } i \in \{3, L\}, \\ \infty & \text{при } 4 \leqslant i \leqslant L-2. \end{cases} \tag{6}$$

Рассмотрим также максимальную устойчивость координат к запретам

$$\mathbf{z}_{\max}(\xi) = \max_{1 \leqslant i \leqslant L} \mathbf{z}(i, \xi).$$

Теперь сформулированное ограничение записывается так: при условии $H(\eta)$ выполнено условие

$$\forall i \notin \Theta \quad (\mathbf{z}(i, \xi) > \mathbf{z}_{\min}(\xi_\Theta)).$$

Оно эквивалентно тому, что $\mathcal{G}_{\min}(\xi)$ не содержит рёбер с номерами из $\bar{\Theta}$, то есть равенству

$$\mathcal{G}_{\min}(\xi) = \mathcal{G}_{\min}(\xi_\Theta). \tag{7}$$

Согласно (6), в примере 2 это условие выполнено.

2. Проверка гипотезы и поиск особенного множества

На основе введённых графов запретов сначала построим критерий согласия с гипотезой $H(\eta)$. В случае априорной справедливости $H(\eta)$ предложен последовательный алгоритм 1 определения $\Theta_{\min}$.

2.1. Критерий на основе графа кратчайших запретов

Предлагается следующий критерий согласия с гипотезой $H(\eta)$:

$$\Big(\mathcal{G}_{\min}(\mathcal{X}(N)) \cong \mathcal{G}_{\min}(\eta)\Big) \quad \Longrightarrow \text{ принимаем гипотезу } H(\eta). \tag{8}$$

Докажем теорему о вероятностях ошибок критерия.

Теорема 1. Пусть $\mathbf{z} = \mathbf{z}_{\min}(\eta) < \infty$. Тогда:

1. При любом $N \geqslant 1$ критерий (8) с вероятностью 1 отклоняет все альтернативы ξ, у которых $\mathbf{z}_{\min}(\xi) < \mathbf{z}$ или $(\mathbf{z}_{\min}(\xi) = \mathbf{z},\ |\mathcal{G}_{\min}(\xi)| > |\mathcal{G}_{\min}(\eta)|)$.
2. При $N \geqslant N_1(\mathbf{z}, \xi, \alpha)$ и альтернативе ξ, такой, что $\mathbf{z}_{\min}(\xi) > \mathbf{z}$, вероятность ошибки критерия не превосходит α.
3. Если выполнено ограничение (7) и $N \geqslant N_1(\mathbf{z}, \xi, \alpha)$, то вероятность ошибки критерия при гипотезе $H(\eta)$ не превосходит α.

Доказательство.

1. Обозначим через A условие-предпосылку в (8), а через $r = \mathbf{z}_{\min}(\mathcal{X})$ — число вершин в рёбрах графа $\mathcal{G}_{\min}(\mathcal{X})$.

Если $r < \mathbf{z}$, то A не выполнено и гипотеза отвергается. С учётом (2), при альтернативе первого вида из п. 1 теоремы имеем

$$r \leqslant \mathbf{z}_{\min}(\xi) < \mathbf{z}.$$

Если $r = \mathbf{z}$ и выполнена альтернатива второго вида, то, согласно (2), граф $\mathcal{G}_{\min}(\mathcal{X})$ содержит граф $\mathcal{G}(\mathbf{z}, \xi)$ и, следовательно, число его рёбер больше числа рёбер $\mathcal{G}(\mathbf{z}, \xi)$. Поэтому здесь также невозможно условие A. Пункт 1 доказан.

2. При этой альтернативе $\mathcal{G}(\mathbf{z}, \xi) = \varnothing$. Поэтому, согласно (4), для $N \geqslant N_1(\mathbf{z}, \xi, \alpha)$ с вероятностью не меньше $1-\alpha$ происходит событие $\mathcal{G}(\mathbf{z}, \mathcal{X}) = \varnothing$, которое несовместно с событием A.

3. Из условий $H(\eta)$ и (7) имеем $\mathcal{G}_{\min}(\eta) \cong \mathcal{G}_{\min}(\xi_\Theta) = \mathcal{G}_{\min}(\xi)$. При $N \geqslant N_1(\mathbf{z}, \xi, \alpha)$, согласно (4), с вероятностью не меньше $1 - \alpha$ происходит событие $\mathcal{G}_{\min}(\xi) = \mathcal{G}_{\min}(\mathcal{X})$, что с учётом предыдущего равенства влечёт событие A. ■

Замечание 1. Особенностью критерия является то, что при его применении не накладывается никаких ограничений на распределение ξ, кроме (7). Но от распределения $\{\mathbf{x}(t)\}_{t\geqslant 1}$ и, в частности, от распределения ξ зависит объём материала $N_1(\cdot)$, достаточный для гарантированной верхней оценки вероятности ошибки критерия.

Замечание 2. Основной вклад в сложность проверки условия (8) может вносить построение графа $\mathcal{G}(\mathbf{z}, \mathcal{X}(N))$, $\mathbf{z} = \mathbf{z}_{\min}(\eta)$. Этот граф можно строить с помощью $\binom{L}{\mathbf{z}}$ битовых массивов длины $k^{\mathbf{z}}$, соответствующих всем $J \in \{1, \ldots, L\}^{(\mathbf{z})}$ и первоначально инициализированных нулями. Проходя по всем наблюдаемым векторам $\mathbf{x}(t)$,

$1 \leqslant t \leqslant N$, будем записывать единицы по адресам $\mathbf{x}(t)_J$ для всех J. После этого нулевые элементы в массиве позволят определить запреты и простые запреты размерности $\mathbf{z}$ эмпирического распределения.

Временная и емкостная сложности такого алгоритма оцениваются величинами порядка $\binom{L}{\mathbf{z}}N$, $\binom{L}{\mathbf{z}}k^{\mathbf{z}}$ соответственно. При больших L они быстро растут с ростом $\mathbf{z}$. Поэтому в критерии согласия для достижения малой временной сложности и для простоты критерия ограничиваемся графами запретов наименьшей размерности. В целом можно предположить, что чем сложнее строение $\mathcal{G}_{\min}(\eta)$, тем мощнее будет критерий (тем больше альтернатив он будет отклонять).

Кратко изложенный подход можно сформулировать так:

1. Для данного η находим $\mathbf{z}$-однородный гиперграф $\mathcal{G}_{\min}(\eta)$, где $\mathbf{z}$ — длина кратчайшего запрета частных распределений η.
2. При статистическом анализе распределения ξ оцениваем гиперграф $\mathcal{G}_{\min}(\xi)$, и если оценка изоморфна графу $\mathcal{G}_{\min}(\eta)$, то принимаем $H(\eta)$.
3. Чем меньше $\mathbf{z}$, тем меньше временная и емкостная сложность проверки критерия, а также величина $N_1(\mathbf{z}, \cdot)$.

2.2. Алгоритмы поиска некоторых номеров координат подвектора с известным либо неизвестным распределением

Везде далее будем считать, что априори справедлива гипотеза $H(\eta)$ и стоит задача определения множества Θ или его части. При известном графе $\mathcal{G}_{\min}(\eta)$ предлагается следующий последовательный алгоритм 1 поиска множества $\Theta_{\min}$.

Алгоритм 1. Поиск множества $\Theta_{\min}$

Вход: $\mathbf{x}(t) \in X^L$, $t = 1, 2, \ldots$, $\quad \mathcal{G}_{\min}(\eta)$
Выход: $\Theta_{\min}$, τ
1: $N := 1$;
2: **Пока** $\mathbf{z}_{\min}(\mathcal{X}(N)) < \mathbf{z}_{\min}(\eta)$ или $|\mathcal{G}_{\min}(\mathcal{X}(N))| > |\mathcal{G}_{\min}(\eta)|$
3: $\quad N := N + 1$;
4: **Вернуть** $\mathbf{V}(\mathcal{G}_{\min}(\mathcal{X}(N)))$ и N

Здесь в качестве статистической оценки $\Theta_{\min}$ рассматривается множество номеров координат, на которых расположены кратчайшие запреты эмпирического распределения.

Теорема 2. Пусть при гипотезе $H(\eta)$ выполнено ограничение (7). Тогда:

1. Вероятность ошибки алгоритма 1 в случае его окончания равна нулю.
2. Для распределения момента τ окончания работы алгоритма 1 справедлива оценка $\mathbf{P}\{\tau > N_1(\mathbf{z}_{\min}(\eta), \xi, \alpha)\} \leqslant \alpha$.

Доказательство.

1. Как и при доказательстве п. 2 теоремы 1, условия $H(\eta)$ и (7) обеспечивают выполнение двух первых соотношений в цепочке

$$\mathcal{G}_{\min}(\eta) \cong \mathcal{G}_{\min}(\xi_\Theta) = \mathcal{G}_{\min}(\xi) \subset \mathcal{G}_{\min}(\mathcal{X}). \tag{9}$$

Из этих условий и первого условия п. 2 алгоритма 1 также следует равенство

$$\mathbf{z}_{\min}(\xi) = \mathbf{z}_{\min}(\eta) = \mathbf{z}_{\min}(\mathcal{X}(N)),$$

которое с учётом (2) даёт последнее включение в (9).

Тогда второе условие п. 2 алгоритма означает равенство количеств рёбер левого и правого графов в цепочке (9). Отсюда следует, что $\mathcal{G}_{\min}(\xi) = \mathcal{G}_{\min}(\mathcal{X})$. Поэтому множества вершин, покрытых рёбрами каждого из них, совпадают.

2. Оценка вероятности следует из (4) и импликаций

$$\big(\mathcal{G}_{\min}(\mathcal{X}) = \mathcal{G}_{\min}(\xi)\big) \Longrightarrow \big(\mathcal{G}_{\min}(\mathcal{X}) \cong \mathcal{G}_{\min}(\eta)\big) \Longrightarrow \begin{pmatrix} \mathbf{z}_{\min}(\mathcal{X}) = \mathbf{z}_{\min}(\eta), \\ |\mathcal{G}_{\min}(\mathcal{X})| = |\mathcal{G}_{\min}(\eta)| \end{pmatrix},$$

справедливых при условиях $H(\eta)$ и (7). ■

Заметим, что если $\Theta_{\min} \neq \Theta$ (что эквивалентно условию $\mathbf{z}_{\min}(\eta) < \mathbf{z}_{\max}(\eta)$), то требуются дополнительные действия для поиска остальных элементов особенного множества.

Далее переходим к алгоритмам поиска части Θ при отсутствии информации о графе $\mathcal{G}_{\min}(\eta)$ и графах запретов большей размерности. В отличие от алгоритма 1, они работают на фиксированном объёме материала, но вероятность их ошибки в общем случае ненулевая.

Алгоритм 2 на фиксированном объёме материала N находит все кратчайшие запреты эмпирического распределения и возвращает множество вершин соответствующего графа в качестве статистической оценки части особенного множества.

Алгоритм 2. Статистическое оценивание $\Theta_{\min}$

Вход: $\mathbf{x}(t) \in X^L$, $1 \leqslant t \leqslant N$
Выход: $\Theta^*_{\min}$
1: **Вернуть** $\mathbf{V}(\mathcal{G}_{\min}(\mathcal{X}(N)))$

Следствие 1. Если при условиях $H(\eta)$ и (7) выполнено $N \geqslant N_1(\mathbf{z}_{\min}(\eta), \xi, \alpha)$, то $\mathbf{P}\{\Theta^*_{\min} = \Theta_{\min}\} \geqslant 1 - \alpha$.

Доказательство. Обозначая $\mathbf{z} = \mathbf{z}_{\min}(\eta)$, из предположения (4) с учётом ограничения (7) получаем, что событие $\{\mathcal{G}(r, \mathcal{X}) = \varnothing,\ 1 \leqslant r < \mathbf{z},\ \mathcal{G}(\mathbf{z}, \mathcal{X}) = \mathcal{G}_{\min}(\xi)\}$ происходит с вероятностью не менее $1 - \alpha$ при $N \geqslant N_1(\mathbf{z}, \xi, \alpha)$. Оно влечёт равенство $\mathcal{G}_{\min}(\mathcal{X}) = \mathcal{G}_{\min}(\xi)$, которое с учётом цепочки (9) доказательства теоремы 2 даёт равенство $\mathcal{G}_{\min}(\mathcal{X}) = \mathcal{G}_{\min}(\xi_\Theta)$. ■

Далее аналогично можем строить статистические оценки множества

$$\Theta_{\leqslant R} := \{i \in \Theta : \mathbf{z}(i, \xi_\Theta) \leqslant R\} \subset \Theta$$

номеров координат, устойчивость к запретам которых не превосходит R, — алгоритм 3.

Алгоритм 3. Построение оценки $\Theta^*_{\leqslant R}$

Вход: $\mathbf{x}(t) \in X^L$, $1 \leqslant t \leqslant N$, $R \geqslant 1$ — параметр алгоритма
Выход: $\Theta^*_{\leqslant R}$
1: **Вернуть** $\Theta^*_{\leqslant R} := \bigcup\limits_{1 \leqslant r \leqslant R} \mathbf{V}(\mathcal{G}(r, \mathcal{X}(N)))$

При некотором ограничении, в общем случае усиливающем ограничение (7), докажем следующую оценку надёжности алгоритма 3.

Следствие 2. Если $N \geqslant N_1(R,\xi,\alpha)$, выполнено условие $H(\eta)$ и ограничение $\forall i \notin \Theta\ (\mathbf{z}(i,\xi) > R)$, то $\mathbf{P}\left\{\Theta^*_{\leqslant R} = \Theta_{\leqslant R}\right\} \geqslant 1-\alpha$.

Доказательство. Согласно ограничению, для каждого $1 \leqslant r \leqslant R$ рёбра графа $\mathcal{G}(r,\xi)$ не содержат номеров из $\bar{\Theta}$, откуда $\Theta_{\leqslant R} = \bigcup\limits_{1\leqslant r\leqslant R} \mathbf{V}(\mathcal{G}(r,\xi))$.

Остаётся заметить, что при $N \geqslant N_1(R,\xi,\alpha)$, согласно предположению (4), событие $\{\mathcal{G}(r,\mathcal{X}) = \mathcal{G}(r,\xi),\ 1 \leqslant r \leqslant R\}$ происходит с вероятностью не менее $1-\alpha$. С учётом предыдущего равенства оно влечёт событие $\left\{\Theta^*_{\leqslant R} = \Theta_{\leqslant R}\right\}$. ■

Не будем останавливаться на оценках сложности методов построения эмпирических простых запретов, а также на оценках сложности проверки условия изоморфизма r-графов и особенностях реализации этих методов — этот круг проблем выходит за рамки работы, причём проблема изоморфизма хорошо известна. Сосредоточимся на вопросах построения графов запретов и оценках числа наблюдений.

3. Расчет параметров критерия и алгоритмов

Для применения критерия (8) согласия с $H(\eta)$ и алгоритмов 1–3 поиска части Θ надо уметь:

1) рассчитывать значение $\mathbf{z}_{\min}(\eta)$, строить граф $\mathcal{G}_{\min}(\eta)$;
2) строить функции $N_1(r,\xi,\alpha)$ и проверять ограничение (7).

Согласно теореме 1, знание функций $N_1(\cdot)$ требуется для оценки надёжности критерия при его применении. Такое же замечание справедливо для алгоритмов 2 и 3, согласно следствиям из теоремы 2.

3.1. Формула для N_1

Получим формулу для оценочной функции $N_1(r,\xi,\alpha)$ в случае независимых наблюдений $\mathbf{x}(t) \sim \xi$. Обозначим далее через

$$p_{\min}(r,\xi) = \min\left\{p = \mathbf{P}\{\xi_{\mathbf{J}} = \mathbf{a}\} : J \in \{1,\ldots,L\}^{(r)},\ \mathbf{a} \in X^r,\ p > 0\right\} > 0$$

самую малую ненулевую вероятность r-мерных распределений ξ.

Теорема 3. Пусть $\mathbf{x}(t) \sim \xi$, $t \geqslant 1$ — последовательность независимых случайных векторов, $0 < \alpha < 1$. Тогда в качестве функции $N_1(\cdot)$ может быть взята функция

$$N_1^*(r,\xi,\alpha) = \frac{1}{p_{\min}(r,\xi)}\left(r\ln\frac{kLe}{r} - \ln\alpha\right).$$

Доказательство. Достаточно доказать, что при $N \geqslant N_1^*(\cdot)$ справедлива оценка (3).

Из неравенства $r! > \left(\frac{r}{e}\right)^r$ имеем $\binom{L}{r} < \frac{L^r}{r!} < \left(\frac{Le}{r}\right)^r$. Положим $p = p_{\min}(r,\xi)$. Используя эту оценку и неравенство $1+x \leqslant e^x$, оцениваем сверху вероятность непоявления какого-либо допустимого подвектора на каких-либо r координатах величиной

$$\mathbf{P}\left\{\exists J \in \{1,\ldots,L\}^{(r)}\ (\mathcal{X}(N)_J \neq \mathrm{Supp}(\xi_J))\right\} \leqslant$$

$$\leqslant \sum_{J\in\{1,\ldots,L\}^{(r)}} \sum_{\mathbf{a}\in\xi_J(\Omega)} \mathbf{P}\left\{\mathbf{x}_J(t) \neq \mathbf{a},\ 1 \leqslant t \leqslant N\right\} \leqslant k^r\binom{L}{r}(1-p)^N < \left(\frac{kLe}{r}\right)^r \exp(-Np),$$

которая не превосходит α при $N \geqslant N_1^*(r,\xi,\alpha)$. ■

Поясним наличие величины $1/p$, $p = p_{\min}(r, \xi)$, в предложенном выражении для N_1. Пусть ν — случайная величина, равная числу наблюдений до первого момента появления исхода с вероятностью p. В схеме независимых наблюдений она имеет геометрическое распределение с параметром p. Для момента τ первого появления всех возможных r-грамм в $\mathcal{X}$ справедлива оценка $\tau \geqslant \nu$. Поэтому для момента остановки τ последовательного алгоритма, ожидающего появления всех возможных r-грамм в $\mathcal{X}$, имеем $\tau \geqslant \nu$. Отсюда получаем $\mathbf{E}\tau \geqslant \mathbf{E}\nu = 1/p$.

Так как

$$\mathbf{P}\{\tau > N\} \geqslant \mathbf{P}\{\nu > N\} = p(1-p)^N + p(1-p)^{N+1} + \ldots = (1-p)^N,$$

вероятность $\mathbf{P}\{\tau > N\}$ может быть не больше α только при $N \geqslant \dfrac{\ln \alpha}{\ln(1-p)}$. Последняя дробь эквивалентна $\dfrac{1}{p} \ln \dfrac{1}{\alpha}$ при $p \to 0$. Поэтому порядок величины $1/p$ в предложенной функции $N_1^*(\cdot)$ не может быть уменьшен.

Теперь можем получить оценку объёма материала, при котором критерий идентификации (8) имеет вероятность ошибки не более заданной величины при гипотезе о равномерности частных распределений ξ. Она очевидно вытекает из теоремы 3 и п. 2 теоремы 1.

Следствие 3. Пусть в схеме независимых наблюдений выполнено

$$\mathbf{z} = \mathbf{z}_{\min}(\eta) < \infty, \quad N \geqslant N_1^*(\mathbf{z}, \xi, \alpha) = 2^{\mathbf{z}} \left(\mathbf{z} \ln \frac{kLe}{\mathbf{z}} - \ln \alpha \right).$$

Тогда при альтернативе о равномерности всех $\mathbf{z}$-мерных частных распределений ξ (сюда, в частности, входит простая гипотеза $\xi \sim U(X^L)$) вероятность ошибки критерия (8) не превосходит $1 - \alpha$.

Получим оценки для $p_{\min}(r, \xi)$, считая по определению $p_{\min}(0, \xi) = 1$.

Теорема 4. Для любого $1 \leqslant r \leqslant \mathbf{z} - 1$, $\mathbf{z} = \mathbf{z}_{\min}(\xi)$, справедливы неравенства

$$p_{\min}(r, \xi) \leqslant \frac{1}{k}\, p_{\min}(r-1, \xi) \leqslant \ldots \leqslant \frac{1}{k^r}\, p_{\min}(0, \xi) = k^{-r}.$$

Следовательно, $p_{\min}(\mathbf{z}, \xi) \leqslant p_{\min}(\mathbf{z} - 1, \xi) \leqslant k^{1-\mathbf{z}}$.

Доказательство. Заметим, что при $r < \mathbf{z}$ каждая вероятность $(r-1)$-мерного распределения ξ равна сумме некоторых k ненулевых вероятностей r-мерных распределений и поэтому не меньше величины $k p_{\min}(r, \xi)$. Отсюда вытекает первое неравенство в первой цепочке, а из него остальные. Во второй цепочке первое неравенство очевидно, а второе получено из первой цепочки при $r = \mathbf{z} - 1$. ■

Таким образом, величина $p_{\min}(r, \xi)$ при $0 \leqslant r \leqslant \mathbf{z} - 1$ убывает от начального значения 1 при каждом увеличении r на 1 не менее чем в k раз; равенство $p_{\min}(r, \xi) = k^{-r}$ достигается тогда и только тогда, когда все r-мерные распределения ξ равномерны. Величина $p_{\min}(\mathbf{z}, \xi)$ может быть как меньше, так и больше величины $k^{-1} p_{\min}(\mathbf{z} - 1, \xi)$.

3.2. Упрощающие условия

Покажем, что условие

$$\text{случайные векторы } \xi_\Theta \text{ и } \xi_{\bar{\Theta}} \text{ независимы} \tag{10}$$

существенно упрощает вычисление $\mathbf{z}_{\min}(\xi)$ и проверку ограничения (7).

Теорема 5. Пусть для некоторого $\varnothing \neq A \subset \{1, \ldots, L\}$ случайные векторы ξ_A и $\xi_{\bar{A}}$ независимы. Тогда

$$\begin{aligned} &1)\ \mathcal{G}(r, \xi) = \mathcal{G}(r, \xi_A) \sqcup \mathcal{G}(r, \xi_{\bar{A}}) \ \text{ для всех } \ 1 \leqslant r \leqslant L; \\ &2)\ \mathbf{z}(a, \xi) = \mathbf{z}(a, \xi_A) \ \text{ для всех } \ a \in A; \\ &3)\ \text{если } \ A = \{a\}, \ \text{ то } \ \mathbf{z}(a, \xi) = \begin{cases} \infty \text{ при } \operatorname{Supp}(\xi_a) = X, \\ 1 \text{ иначе.} \end{cases} \end{aligned} \tag{11}$$

Доказательство.

1. Предположим противное: пусть существует $\mathbf{b}$ — простой запрет распределения $\xi_{I \cup J}$, где $\varnothing \neq I \subset A$; $\varnothing \neq J \subset \bar{A}$. Тогда из условия независимости имеем

$$0 = \mathbf{P}\{\xi_{I \cup J} = \mathbf{b}\} = \mathbf{P}\{\xi_I = \mathbf{b}_I\}\mathbf{P}\{\xi_J = \mathbf{b}_J\},$$

и один из сомножителей в последнем выражении равен нулю. Но это противоречит простоте запрета. Пункт 1 доказан.

2. Пункт 2 следует из п. 1, поскольку вершина $a \in A$ может лежать только в рёбрах графов $\mathcal{G}(r, \xi_A)$.

3. Если $\operatorname{Supp}(\xi_a) \neq X$, то утверждение очевидно. В противном случае распределение ξ_a не имеет запретов, а в r-мерных простых запретах при $r \geqslant 2$ номер a не участвует согласно п. 1. Пункт 3 доказан. ■

Равенство (11) означает, что если ξ можно разделить на два независимых подвектора, то графы запретов распадаются на не связанные между собою графы запретов подвекторов. Можно также показать, что условие независимости в теореме 5 не является необходимым для условия (11).

Следствие 4. Если при справедливости гипотезы $H(\eta)$ выполнено условие (10), то ограничение (7) равносильно условию $\mathbf{z}_{\min}(\eta) < \mathbf{z}_{\min}(\xi_{\bar{\Theta}})$.

Доказательство. Из п. 2 теоремы 5 имеем $\min\limits_{i \in \Theta} \mathbf{z}(i, \xi) = \mathbf{z}_{\min}(\xi_\Theta)$, $\min\limits_{i \in \bar{\Theta}} \mathbf{z}(i, \xi) = \mathbf{z}_{\min}(\xi_{\bar{\Theta}})$. Осталось заметить, что при гипотезе $H(\eta)$ выполнено равенство $\mathbf{z}_{\min}(\xi_\Theta) = \mathbf{z}_{\min}(\eta)$. ■

Введём второе упрощающее условие

$$\xi_\Theta \ \text{ не зависит от } \ \xi_{\bar{\Theta}} \sim U(X^{L - |\Theta|}). \tag{12}$$

Оно получено путём добавления в условие независимости (10) условия равномерности распределения координат, не принадлежащих подвектору. Можно сказать, что в этом случае подвектор на координатах с номерами из Θ «погружен» в не зависящий от него равномерно распределённый случайный вектор $\xi_{\bar{\Theta}}$.

При условии (12) из п. 3 теоремы 5 получаем, что $\mathbf{z}_{\min}(\xi_{\bar{\Theta}}) = \infty$. Тогда, согласно следствию 4, выполнено ограничение (7) и корректно далее используемое единое обозначение для устойчивостей к запретам трёх распределений

$$\mathbf{z} = \mathbf{z}_{\min}(\xi) = \mathbf{z}_{\min}(\xi_\Theta) = \mathbf{z}_{\min}(\eta).$$

Кроме того, при условии (12) для любого R выполнено ограничение следствия 2, при котором оценивалась надёжность алгоритма 3.

При введённом условии (12) и гипотезе $H(\eta)$ распределение ξ определяется распределением η (с точностью до перестановки координат). Поэтому можем явно выразить фигурирующую в оценках объёма материала величину $p_{\min}(\mathbf{z}, \xi)$ через распределение η.

Теорема 6. Если выполнены условия $H(\eta)$, (12) и $\mathbf{z} < L$, то

$$p_{\min}(r,\xi) = \min\{p_{\min}(r,\eta), \frac{1}{k}\, p_{\min}(r-1,\eta)\},\ 1 \leqslant r \leqslant \mathbf{z}; \tag{13}$$

$$p_{\min}(r,\xi) \leqslant k^{-r},\ 1 \leqslant r \leqslant \mathbf{z}. \tag{14}$$

Доказательство. Докажем утверждения в случае $r = 1$. Здесь имеем

$$p_{\min}(1,\xi) = \min\{p_{\min}(1,\xi_\Theta), p_{\min}(1,\xi_{\bar\Theta})\},$$

что совпадает с правой частью (13), поскольку $p_{\min}(1,\xi_{\bar\Theta}) = \dfrac{1}{k} = \dfrac{1}{k}\, p_{\min}(0,\eta)$. Из последней формулы также вытекает справедливость (14) при $r = 1$.

Осталось доказать утверждения в случае $2 \leqslant r \leqslant L-1$, $\mathbf{z} \geqslant 2$. Очевидно, что для всех $2 \leqslant r \leqslant L-1$ выполнено равенство

$$p_{\min}(r,\xi) = \min\{p_{\min}(r,\xi_\Theta), p_{\min}(r,\xi_{\bar\Theta}), p_{\mathrm{joi}}(r)\}, \tag{15}$$

где $p_{\mathrm{joi}}(r)$ — минимум ненулевых вероятностей r-мерных распределений на координатах, содержащих номера из Θ и $\bar\Theta$ одновременно.

Используя равномерность распределения $\xi_{\bar\Theta}$ в первом переходе и первую цепочку неравенств теоремы 4 во втором переходе, при $2 \leqslant r \leqslant \mathbf{z}$ имеем равенства

$$p_{\mathrm{joi}}(r) = \min_{1\leqslant l\leqslant r-1} p_{\min}(l,\xi_\Theta)k^{l-r} = p_{\min}(r-1,\xi_\Theta)k^{-1} = p_{\min}(r-1,\eta)k^{-1}.$$

Согласно (15) и полученному выражению для $p_{\mathrm{joi}}(\mathbf{z})$, для доказательства равенства (13) осталось обосновать неравенство

$$p_{\min}(r-1,\eta)k^{-1} \leqslant p_{\min}(r,\xi_{\bar\Theta}) = k^{-r}.$$

Но при всех $2 \leqslant r \leqslant \mathbf{z}$ оно следует из первой цепочки теоремы 4.

Из последнего неравенства также вытекает (14). ■

Заметим, что при $\mathbf{z} = L$ утверждения теоремы могут быть неверны. В частности, тогда в (13) левая часть равна первому выражению под знаком минимума, а второе выражение под знаком минимума может быть меньше первого. Например, для $\xi = \eta \equiv (0)$ имеем $p_{\min}(1,\xi) = 1 > \dfrac{1}{k} = \dfrac{1}{k}\, p_{\min}(0,\eta)$.

О возможности обобщения результатов на случай $r > \mathbf{z}$ заметим следующее. Равенство (13) прямого обобщения не допускает, поскольку, например, при $\mathbf{z} = M$ величина $p_{\min}(r,\eta)$ не определена для $r > \mathbf{z}$. Неравенство (14) при $r > \mathbf{z}$ в общем случае неверно. Например, при $\xi = (0,x)$, $x \sim U(X)$ имеем $\mathbf{z} = 1$, $p_{\min}(2,\xi) = \dfrac{1}{k} > k^{-2}$.

Итак, при гипотезе $H(\eta)$ дополнительное условие (12) позволяет по величинам, полностью определяемым распределением η:

— построить критерий и алгоритм 1;

— рассчитать оценку $N_1^*(\mathbf{z}_{\min}(\eta),\xi,\alpha)$ объёма материала критерия и алгоритмов 1 и 2.

Заметим, что при этом условии из (14) вытекает нижняя оценка для достаточного числа наблюдений: для всех $1 \leqslant r \leqslant \mathbf{z}_{\min}(\eta)$

$$N_1^*(r,\xi,\alpha) \geqslant k^r \left(r\left(\ln\frac{kL}{r} + 1\right) - \ln\alpha \right).$$

Заключение

Определение запрета размерности r на множестве номеров координат $J \in \in \{1,\dots,L\}^{(r)}$, приведённое в работе, обобщает определение 1 [2, с. 55], в котором такие запреты рассматриваются лишь при $J = \{1,\dots,r\}$. Поэтому наше понятие кратчайшего запрета отличается от понятий наименьшего и минимального запретов [2, с. 55]. В наших терминах $\mathbf{a} \in X^s$ является минимальным запретом в смысле [2], если он является запретом на множестве номеров координат $J = \{n-s+1,\dots,n\}$ (последних s координатах) и не является запретом на множестве $\{n-s+2,\dots,n\}$.

В [2, с. 55] предложен алгоритм построения статистической оценки множества всех минимальных запретов распределения ξ размерности не более s_0 при независимых наблюдениях. Получено выражение для математического ожидания числа эмпирических запретов (на последних координатах) размерности не более s_0, а также верхняя оценка для математического ожидания числа $\nu(s_0)$ таких эмпирических запретов, не являющихся теоретическими:

$$\mathbf{E}\nu(s_0) \leqslant s_0 k^{s_0}(1 - p'_{\min}(s_0,\xi))^N,$$

где $p'_{\min}(s_0,\xi)$ — минимум ненулевых вероятностей распределения на последних s_0 координатах. Очевидно, что эта величина связана с аналогичной введённой величиной неравенством $p'_{\min}(s_0,\xi) \geqslant p_{\min}(s_0,\xi)$. Из неравенства для $\mathbf{E}\nu(s_0)$ с помощью неравенства Маркова сделан вывод о состоятельности при $N \to \infty$ предложенной оценки множества таких запретов.

Ранее в [5] и других работах А. А. Грушо и Е. Е. Тимониной исследовалась задача построения состоятельных критериев, которые предполагалось применять для статистического выявления сбоев в протоколах и технических устройствах.

Проведённая конкретизация задачи и развитие понятийного аппарата позволили сформулировать конкретные алгоритмы её решения, дать допредельные оценки вероятностей ошибок. Такое направление развития предполагалось в заключении работы [1].

Предложенный подход, в свою очередь, может развиваться в следующих направлениях:

— получение явных верхних оценок числа наблюдений для критерия и алгоритмов в случае «функциональной вставки» $\eta \sim \mathbf{xf}$, в том числе для конкретных классов функций;
— получение оценок числа наблюдений для более сложных распределений выборки, например схемы конечно зависимых наблюдений, схемы псевдослучайного образования аргументов функции.

ЛИТЕРАТУРА

1. *Грушо А. А., Тимонина Е. Е.* Запреты в дискретных вероятностно-статистических задачах // Дискретная математика. 2011. Т. 23. № 2. С. 53–58.
2. *Грушо А. А., Грушо Н. А., Тимонина Е. Е.* Статистические методы определения запретов вероятностных мер на дискретных пространствах // Информатика и её применение. 2013. Т. 7. № 1. С. 54–57.
3. *Михайлов В. Г., Чистяков В. П.* О задачах теории конечных автоматов, связанных с числом прообразов выходной последовательности // Обозрение прикл. и промышл. матем. Сер. дискретн. матем. 1994. Т. 1. Вып. 1. С. 7–32.

4. *Сумароков С. Н.* Запреты двоичных функций и обратимость для одного класса кодирующих устройств // Обозрение прикл. и промышл. матем. Сер. дискретн. матем. 1994. Т. 1. Вып. 1. С. 33–55.
5. *Грушо А. А., Тимонина Е. Е.* Некоторые связи между дискретными статистическими задачами и свойствами вероятностных мер на топологических пространствах // Дискретная математика. 2006. Т. 18. № 4. С. 128–136.

REFERENCES

1. *Grusho A. A., Timonina E. E.* Zaprety v diskretnykh veroyatnostno-statisticheskikh zadachakh [Prohibitions in discrete probabilistic statistical problems.] Diskretnaya Matematika, 2011, vol. 23, no. 2, pp. 53–58. (in Russian)
2. *Grusho A. A., Grusho N. A., Timonina E. E.* Statisticheskie metody opredeleniya zapretov veroyatnostnykh mer na diskretnykh prostranstvakh [Statistical techniques of bans determination of probability measures in discrete spaces.] Inform. Primen., 2013, vol. 7, no. 1, pp. 54–57. (in Russian)
3. *Mikhaylov V. G., Chistyakov V. P.* O zadachakh teorii konechnykh avtomatov, svyazannykh s chislom proobrazov vykhodnoy posledovatel'nosti [Problems of the finite automata theory associated with a number of inverse images of the output sequence.] Obozrenie Prikl. i Promyshl. Matem. Ser. Diskretn. Matem., 1994, vol. 1, iss. 1, pp. 7–32. (in Russian)
4. *Sumarokov S. N.* Zaprety dvoichnykh funktsiy i obratimost' dlya odnogo klassa kodiruyushchikh ustroystv [Prohibitions of binary functions and reversibility for a class of encoders.] Obozrenie Prikl. i Promyshl. Matem. Ser. Diskretn. Matem., 1994, vol. 1, iss. 1, pp. 33–55. (in Russian)
5. *Grusho A. A., Timonina E. E.* Nekotorye svyazi mezhdu diskretnymi statisticheskimi zadachami i svoystvami veroyatnostnykh mer na topologicheskikh prostranstvakh [Some relations between discrete statistical problems and properties of probability measures on topological spaces.] Diskretnaya Matematika, 2006, vol. 18, no. 4, pp. 128–136. (in Russian)