

УДК 621.396

DOI: [10.26102/2310-6018/2020.31.4.034](https://doi.org/10.26102/2310-6018/2020.31.4.034)

## Энтропийные оценки решающих статистик алгоритма классификации случайных процессов

М.Ю. Калинин<sup>1</sup>, О.Н. Чопоров<sup>2</sup>

<sup>1</sup> ООО «Импульс-сервис», Москва, Российская Федерация

<sup>2</sup> АНОО ВО Воронежский институт высоких технологий,  
Воронеж, Российская Федерация

**Резюме:** Рассматривается задача выявления состояния технической системы по поступающим от нее сигналам, каждому состоянию соответствует класс сигналов с определенными свойствами, актуальная в области распознавания образов, технической диагностики и других направлениях науки и техники. Для ее решения определяется принадлежность поступающего сигнала одному из выбранных классов. Для описания случайного сигнала и математического представления классов используется марковская модель случайного процесса, на основе которой разработан оптимальный алгоритм классификации сигнала с заданной достоверностью. Получены величины (решающие статистики), по которым принимается решение о принадлежности выборки отсчетов принимаемого сигнала соответствующему классу и которые позволяют оценивать «расстояние» между классами (их моделями). Их изучение позволяет оценивать возможности и эффективность алгоритмов классификации сигналов, а также свойства множества классов по их марковским моделям. С использованием теории информации исследуются свойства решающих статистик, определяются их вероятностные характеристики. С использованием понятий энтропии и информационной дивергенции (расстояния Кульбака-Лейблера) получены оценки среднего значения и дисперсии решающих статистик. Получены оценки продолжительности процедуры классификации. Приведен пример расчета. Результаты исследований могут быть использованы при определении состояния технических устройств (двигателей, турбин и др.) по поступающим сигналам от размещенных на них датчиков, при классификации радиосигналов в системах радиомониторинга и других научных и технических приложениях.

**Ключевые слова:** сигнал, классификация, марковская модель, энтропия, информационная дивергенция.

**Для цитирования:** Калинин М.Ю., Чопоров О.Н. Энтропийные оценки решающих статистик алгоритма классификации случайных процессов. *Моделирование, оптимизация и информационные технологии*. 2020;8(4). Доступно по: <https://moitvvt.ru/ru/journal/pdf?id=881>  
DOI: 10.26102/2310-6018/2020.31.4.034

## Entropy estimates of the decision statistics of the classification algorithm for random processes

M.Y. Kalinin<sup>1</sup>, O.N. Choporov<sup>2</sup>

<sup>1</sup> Impulse-Service, Moscow, Russia

<sup>2</sup> Voronezh institute of high technologies, Voronezh, Russian Federation

**Abstract:** The problem of identifying the state of a technical system based on signals coming from it is considered, each state corresponds to a class of signals with certain properties, which is relevant in the field of pattern recognition, technical diagnostics and other areas of science and technology. To solve it, the belonging of the incoming signal to one of the selected classes is determined. To describe a random signal and mathematical representation of classes, a Markov model of a random process is used,

on the basis of which an optimal signal classification algorithm with a given reliability has been developed. Values (decisive statistics) are obtained, according to which a decision is made about the belonging of a sample of samples of the received signal to the corresponding class and which allow us to estimate the "distance" between the classes (their models). Their study allows one to evaluate the capabilities and efficiency of signal classification algorithms, as well as the properties of a set of classes by their Markov models. With the use of information theory, the properties of decisive statistics are investigated, their probabilistic characteristics are determined. Using the concepts of entropy and information divergence (Kullback-Leibler distance), estimates of the mean value and variance of the decision statistics are obtained. Estimates of the duration of the classification procedure are obtained. An example of calculation is given. The research results can be used to determine the state of technical devices (engines, turbines, etc.) based on incoming signals from sensors placed on them, when classifying radio signals in radio monitoring systems and other scientific and technical applications.

**Keywords:** signal, classification, Markov model, entropy, information divergence

**For citation:** Kalinin M.Y., Choporov O.N. Entropy estimates of the decision statistics of the classification algorithm for random processes. *Modeling, optimization and information technology*. 2020;8(4). Available from: <https://moitvvt.ru/ru/journal/pdf?id=881> DOI: 10.26102/2310-6018/2020.31.4.034 (In Russ).

### Введение (Introduction)

В различных областях техники возникает необходимость выявления состояния объекта наблюдения по поступающим от него информационным случайным сигналам, то есть определения принадлежности наблюдаемого сигнала к одному из нескольких классов, соответствующих состоянию анализируемого объекта. Примерами могут служить задачи определения состояния механического устройства по его вибрациям или выявления типа принимаемого радиосигнала. Подобные проблемы решаются в теории информации [1], распознавания образов [2], кластеризации [3], статистических решений [4] и других направлениях науки и техники.

Ключевой проблемой в этих областях является выбор математической модели анализируемых случайных сигналов. Простая модель (например, гауссовского случайного процесса) упрощает анализ, но ограничивает применимость и общность полученных результатов, сложная модель усложняет решение задачи классификации. Компромиссным является выбор марковской модели [5, 6] случайных процессов (цепи Маркова), обладающей достаточной общностью их описания и позволяющей реализовать оптимальный алгоритм обработки отсчетов сигналов для решения задачи их классификации и исследовать его свойства [7].

### Материалы и методы (Materials and Methods)

В цифровых системах обработки сигналов они дискретизируются по времени и уровню, образуя последовательность отсчетов  $z_n$  ( $n$  – его номер) с целочисленными значениями от 1 до  $M$ . Совокупность отсчетов  $z_n, n = \overline{1, N}$  называют выборкой,  $N$  – ее объем.

Для простой цепи Маркова [6] значение отсчета  $z_n = j$  зависит только от значения  $z_{n-1} = i$  и возникает с переходной вероятностью  $P_{ij}$  и модель описывается стохастической матрицей переходных вероятностей

$$P_{ij} = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1M} \\ P_{21} & P_{22} & \dots & P_{2M} \\ \cdot & \cdot & \cdot & \cdot \\ P_{M1} & P_{M2} & \dots & P_{MM} \end{bmatrix} \quad (1)$$

с условием нормировки

$$\sum_{j=1}^M P_{ij} = 1, \quad i = \overline{1, M}, \quad (2)$$

Начальное значение отсчета  $z_1 = i$  появляется с вероятностью  $P_{z_1}$ .

Марковская модель может быть построена по известным теоретическим статистическим свойствам моделируемых объектов в виде

$$P_{ij} = \frac{\int_{g_{i-1}}^{g_i} \int_{g_{j-1}}^{g_j} w(x_1, x_2) dx_2 dx_1}{\int_{g_{i-1}}^{g_i} \int_{-\infty}^{\infty} w(x_1, x_2) dx_2 dx_1}, \quad (3)$$

где  $w(x_1, x_2)$  – двумерная плотность вероятностей непрерывных значений сигнала в соседние моменты времени дискретизации, а  $g_i$  – пороги квантования сигнала по уровню в аналого-цифровом преобразователе (АЦП),  $i = \overline{1, M}$ .

Модель может быть сформирована по экспериментальным реализациям отсчетов сигнала от объекта в известном состоянии в виде статистической оценки переходных вероятностей

$$\tilde{P}_{ij} = \frac{l_{ij}}{\sum_{k=1}^M l_{ik}}, \quad (4)$$

где  $l_{ij}$  – число переходов отсчетов от предшествующего значения  $z_{n-1} = i$  к текущему  $z_n = j$ . При большом объеме выборки значения  $\tilde{P}_{ij}$  близки к  $P_{ij}$ .

Для сигналов каждого  $k$ -го класса (состояния объекта)  $G_k$  строятся соответствующие модели  $P_{ij}^{(k)}$ ,  $k = \overline{1, K}$ ,  $K$  число классов, с которыми сравнивается выборка отсчетов наблюдаемого сигнала  $z_n, n = \overline{1, N}$ , вычисляются оценки ее близости к марковским моделям классов и на этой основе формируется классификационное решение.

Для выборки из  $N$  отсчетов  $z_1, z_2, \dots, z_N$  наблюдаемого сигнала  $s(t)$  апостериорная вероятность ее принадлежности классу  $G_k$  равна

$$P(G_k / z_1, z_2, \dots, z_N) = \frac{P(G_k) \cdot P(z_1, z_2, \dots, z_N / G_k)}{P(z_1, z_2, \dots, z_N)}, \quad (5)$$

где  $P(G_k)$  — априорная вероятность появления из  $k$ -го класса,  $P(z_1, z_2, \dots, z_N / G_k)$  — условная вероятность возникновения наблюдаемой выборки, если она принадлежит классу  $G_k$ , а  $P(z_1, z_2, \dots, z_N)$  — безусловная вероятность ее возникновения. На основе свойств марковской модели, обозначив

$$L_k = L_k(z_1, z_2, \dots, z_N) = -\sum_{i=1}^M \sum_{j=1}^M l_{ij} \cdot \log_2 P_{ij}^{(k)}, \quad (6)$$

можно показать, что

$$P(z_1, z_2, \dots, z_N / G_k) = p_{z_1}^{(k)} \cdot 2^{-L_k(z_1, z_2, \dots, z_N)}. \quad (7)$$

Аналогично для безусловной вероятности выборки получим

$$\begin{aligned} P(z_1, z_2, \dots, z_N) &= \sum_{m=1}^K P(G_m) \cdot P(z_1, z_2, \dots, z_N / G_m) = \\ &= \sum_{m=1}^K p_{z_1}^{(m)} \cdot P(G_m) \cdot 2^{-L_m(z_1, z_2, \dots, z_N)}. \end{aligned} \quad (8)$$

тогда из (5) апостериорная вероятность принадлежности наблюдаемой выборки классу  $G_k$  равна

$$P(G_k / z_1, z_2, \dots, z_N) = \frac{p_{z_1}^{(k)} \cdot P(G_k) \cdot 2^{-L_k(z_1, z_2, \dots, z_N)}}{\sum_{m=1}^K p_{z_1}^{(m)} \cdot P(G_m) \cdot 2^{-L_m(z_1, z_2, \dots, z_N)}}. \quad (9)$$

В теории статистических решений [1-4] чаще всего используется критерий максимума апостериорной вероятности, тогда из (9) оптимальный алгоритм классификации имеет вид

$$P(G_{k_0} / z_1, z_2, \dots, z_N) = \max_{k \neq k_0} (P(G_k / z_1, z_2, \dots, z_N)), \quad (10)$$

то есть принимается решение о принадлежности выборки классу  $k_0$  с наибольшей апостериорной вероятностью. Тогда из (9) следует, что необходимо выбрать класс  $G_{k_0}$ , для которого величина  $L_{k_0} = L_{k_0}(z_1, z_2, \dots, z_N)$  минимальна,

$$L_k(z_1, z_2, \dots, z_N) - L_{k_0}(z_1, z_2, \dots, z_N) \geq 0, \quad k = \overline{1, K}, \quad k \neq k_0. \quad (11)$$

и тогда величины  $L_k$  целесообразно назвать решающими статистиками.

Алгоритм (11) выбирает наиболее вероятное решение, но не определяет его достоверность. Если потребовать, чтобы максимальная апостериорная вероятность (9) была не меньше заданного значения  $P_0$ , то оптимальный алгоритм формирования решения имеет вид

$$L_{k_1}(z_1, z_2, \dots, z_N) - L_{k_0}(z_1, z_2, \dots, z_N) \geq D, \quad (12)$$

где

$$D = \log_2 \left[ \frac{\sum_{k=1, k \neq k_0}^K p_{z_1}^{(k)} \cdot P(G_k)}{p_{z_1}^{(k_0)} P(G_{k_0})} \frac{P_0}{1 - P_0} \right], \quad (13)$$

где  $L_{k_1}(z_1, z_2, \dots, z_N)$  - минимальное значение решающих статистик за исключением  $L_{k_0}(z_1, z_2, \dots, z_N)$ , а  $D$  - порог решения, при котором обеспечивается требуемая достоверность  $P_0$  выбора класса  $G_{k_0}$ . При равновероятных начальных значениях выборки и классах из (13) получим

$$D = \log_2 \left[ (K - 1) \frac{P_0}{1 - P_0} \right]. \quad (14)$$

Таким образом, оптимальный алгоритм классификации требует превышения разности минимальных решающих статистик порога  $D$ , определяемого требуемой

достоверностью решения.

Значения решающей статистики (6) случайны, так как зависит от случайных чисел  $l_{ij}$ , и могут быть записаны в виде

$$L_k = -(N-1) \cdot \sum_{i=1}^M \sum_{j=1}^M \frac{l_{ij}}{N-1} \cdot \log_2 P_{ij}^{(k)}. \quad (15)$$

При  $N \gg M^2$  величина  $l_{ij}/(N-1)$  является оценкой совместной вероятности  $Q(i, j)$  появления пар отсчетов  $z_{n-1} = i$  и  $z_n = j$ ,

$$Q(i, j) = \frac{l_{ij}}{N-1}, \quad (16)$$

тогда

$$L_k = -(N-1) \cdot \sum_{i=1}^M \sum_{j=1}^M Q(i, j) \cdot \log_2 P_{ij}^{(k)}. \quad (17)$$

Переходные вероятности можно представить в виде

$$P_{ij}^{(k)} = \frac{P^{(k)}(i, j)}{P^{(k)}(i)}, \quad (18)$$

где  $P^{(k)}(i, j)$  — совместное распределение вероятностей значений  $z_{n-1} = i, z_n = j$  в классе  $G_k$ ,  $P^{(k)}(i)$  — одномерное распределение вероятностей значений отсчетов в классе  $G_k$ , тогда из (17) получим

$$\begin{aligned} L_k &= -(N-1) \cdot \sum_{i=1}^M \sum_{j=1}^M Q(i, j) \cdot \log_2 [P^{(k)}(i, j)] + (N-1) \cdot \sum_{i=1}^M \sum_{j=1}^M Q(i, j) \cdot \log_2 [P^{(k)}(i)] = \\ &= -(N-1) \cdot \sum_{i=1}^M \sum_{j=1}^M Q(i, j) \cdot \log_2 [P^{(k)}(i, j)] + (N-1) \cdot \sum_{i=1}^M Q(i) \cdot \log_2 [P^{(k)}(i)], \end{aligned} \quad (19)$$

где

$$Q(i) = \sum_{j=1}^M Q(i, j) \quad (20)$$

— одномерная вероятность появления в выборке значения  $z = i$ .

Тогда для решающей статистики получим

$$L_k = (N-1) [H_2(Q, P^{(k)}) - H_1(Q, P^{(k)})], \quad (21)$$

где

$$H_2(Q, P^{(k)}) = -\sum_{i=1}^M \sum_{j=1}^M Q(i, j) \cdot \log_2 [P^{(k)}(i, j)] \quad (23)$$

— это перекрестная энтропия [1] процессов с двумерными распределениями  $Q(i, j)$  и  $P^{(k)}(i, j)$ , а

$$H_1(Q, P^{(k)}) = -\sum_{i=1}^M Q(i) \cdot \log_2 [P^{(k)}(i)] \quad (24)$$

— перекрестная энтропия процессов с одномерными распределениями  $Q(i)$  и  $P^{(k)}(i)$ . Таким образом, среднее значение решающей статистики  $L_k$  пропорционально объему выборки  $N$ ,

$$L_k = (N-1) \cdot H(Q, P^{(k)}), \quad (25)$$

с коэффициентом  $H(Q, P^{(k)})$ , равным

$$H(Q, P^{(k)}) = H_2(Q, P^{(k)}) - H_1(Q, P^{(k)}). \quad (26)$$

В теории информации [8] рассматривается информационная дивергенция, которая распределений вероятностей  $p_i$  и  $q_i$ ,  $i = \overline{1, M}$  представляется в виде

$$D(p, q) = \sum_{i=1}^M p_i \log_2 \left( \frac{p_i}{q_i} \right) \quad (27)$$

и характеризует взаимную различимость (удаленность) этих распределений вероятностей. Ее также называют расстоянием Кульбака-Лейблера [9].

С учетом (27) получим

$$H(Q, P^{(k)}) = H_2(Q) - H_1(Q) + D_2(Q, P^{(k)}) - D_1(Q, P^{(k)}), \quad (28)$$

где двумерная и одномерная энтропии сигнала равны

$$H_2(Q) = -\sum_{i=1}^M \sum_{j=1}^M Q(i, j) \cdot \log_2 [Q(i, j)], \quad (29)$$

$$H_1(Q) = -\sum_{i=1}^M Q(i) \cdot \log_2 [Q(i)], \quad (30)$$

а двумерная и одномерная информационные дивергенции соответственно

$$D_2(Q, P^{(k)}) = -\sum_{i=1}^M \sum_{j=1}^M Q(i, j) \cdot \log_2 \left[ \frac{Q(i, j)}{P^{(k)}(i, j)} \right], \quad (31)$$

$$D_1(Q, P^{(k)}) = -\sum_{i=1}^M Q(i) \cdot \log_2 \left[ \frac{Q(i)}{P^{(k)}(i)} \right]. \quad (32)$$

Таким образом среднее значение решающей статистики пропорционально разности двумерных энтропий и информационных дивергенций наблюдаемого сигнала.

Для выборки из класса  $G_k$  при  $N \gg 1$  приближенно  $Q(i, j) \approx P^{(k)}(i, j)$  и  $Q(i) \approx P^{(k)}(i)$ , при этом информационные дивергенции в (28) равны нулю и для решающей статистики получим

$$L_k = (N - 1) \cdot [H_2(P^{(k)}(i, j)) - H_1(P^{(k)}(i))], \quad (33)$$

где  $P^{(k)}(i, j)$  — совместное распределение вероятностей соседних пар отсчетов в классе  $G_k$ , а  $P^{(k)}(i)$  — их одномерное распределение вероятностей.

Из (33) следует, что среднее значение решающей статистики для выборки из класса  $G_k$  пропорционально разности двумерной и одномерной энтропий отсчетов в этом классе.

Если соседние отсчеты независимы, то марковская модель описывается одномерным распределением вероятностей и

$$H_2(P^{(k)}(i, j)) = 2H_1(P^{(k)}(i)), \quad (34)$$

тогда

$$L_k = (N - 1) \cdot H_1(P^{(k)}(i)). \quad (35)$$

Решающая статистика  $L_k$  (6) является взвешенной суммой случайных чисел перехода  $l_{ij}$ , тогда в соответствии с центральной предельной теоремой значения  $L_k$  имеют нормальную плотность вероятностей со средним (33) и согласно [10, 11] для дисперсии  $\sigma^2(L_k) = \sigma_k^2$  решающей статистики получим

$$\sigma_k^2 = (N-1) \sum_{i=1}^M \sum_{j=1}^M Q(i, j) \left( \log_2 P_{ij}^{(k)} \right)^2 + C(L_k), \quad (36)$$

$$C(L_k) = \sum_{i_1=1}^M \sum_{j_1=1}^M \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^M \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^M \log_2 P_{i_1 j_1}^{(k)} \cdot \log_2 P_{i_2 j_2}^{(k)} \cdot \text{cov}(l_{i_1 j_1}, l_{i_2 j_2}). \quad (37)$$

Первая двойная сумма в (36) определяется собственной вариацией чисел перехода  $l_{ij}$  в наблюдаемой выборке, а  $C(L_k)$  зависит от их взаимной корреляции. Так как

$$\sum_{i=1}^M \sum_{j=1}^M l_{ij} = N-1, \quad (38)$$

то значения  $l_{ij}$  обладают отрицательной ковариацией, тогда  $C(L_k) < 0$  и для дисперсии  $\sigma^2(L_k)$  получим верхнюю границу в виде

$$\sigma_k^2 \leq (N-1) \sum_{i=1}^M \sum_{j=1}^M Q(i, j) \left( \log_2 P_{ij}^{(k)} \right)^2, \quad (39)$$

а для среднеквадратического отклонения  $\sigma_k$  соответственно

$$\sigma_k = \sqrt{\sigma^2(L_k)} \leq \sqrt{(N-1)} \sqrt{\sum_{i=1}^M \sum_{j=1}^M Q(i, j) \left( \log_2 P_{ij}^{(k)} \right)^2}. \quad (40)$$

Согласно (12) алгоритм классификации предусматривает вычисление разности минимальных решающих статистик  $\Delta L = l_{k1} - l_{k0}$  и сравнение ее с заданным порогом  $D$ . Среднее значение  $\Delta L$  равно

$$\Delta L = (N-1) \cdot \left\{ \left[ D_2(Q, P^{(k1)}) - D_1(Q, P^{(k1)}) \right] - \left[ D_2(Q, P^{(k0)}) - D_1(Q, P^{(k0)}) \right] \right\} \quad (41)$$

и, как видно, определяется информационными дивергенциями первого и второго порядков соответствующих классов.

Для дисперсии  $\sigma^2(\Delta L)$  и среднеквадратического отклонения  $\sigma_{\Delta L}$  разности  $\Delta L$  получим

$$\sigma^2(\Delta L) \leq (N-1) \sum_{i=1}^M \sum_{j=1}^M Q(i, j) \left[ \left( \log_2 P_{ij}^{(k1)} \right)^2 + \left( \log_2 P_{ij}^{(k)} \right)^2 \right], \quad (42)$$

$$\sigma_{\Delta L} = \sqrt{N-1} \sqrt{\sum_{i=1}^M \sum_{j=1}^M Q(i, j) \left[ \left( \log_2 P_{ij}^{(k1)} \right)^2 + \left( \log_2 P_{ij}^{(k)} \right)^2 \right]}, \quad (43)$$

таким образом среднее значение  $\Delta L$  растет пропорционально  $N$ , а  $\sigma_{\Delta L}$  как  $\sqrt{N}$ , то есть значительно медленнее.

### Результаты (Results)

Рассмотрим свойства решающих статистик алгоритма классификации некоррелированных нормальных случайных процессов с нулевым средним для двух классов ( $K=2$ ) со среднеквадратическими отклонениями  $\sigma_1=1$  (класс  $G_1$ ) и  $\sigma_2=1,5$  (класс  $G_2$ ), трехмерные диаграммы соответствующих матриц переходных вероятностей приведены на Рисунке 1.

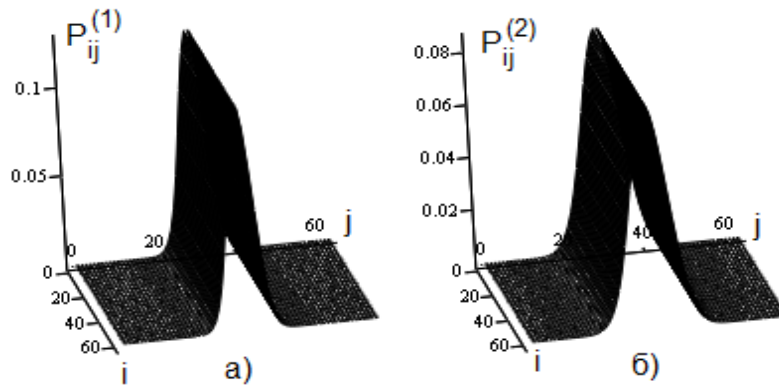


Рисунок 1 – Трехмерные диаграммы матриц переходных вероятностей  
 Figure 1 – Three-dimensional diagrams of matrices of transition probabilities

Пусть поступают выборки отсчетов при  $M=64$  объемом  $N = 100$ , относящиеся к классу  $G_1$  ( $k_0 = 1$ ), тогда решающие статистики равны

$$L_1 = -\sum_{i=1}^M \sum_{j=1}^M l_{ij}^{(1)} \cdot \log_2 P_{ij}^{(1)}, \quad (44)$$

$$L_2 = -\sum_{i=1}^M \sum_{j=1}^M l_{ij}^{(2)} \cdot \log_2 P_{ij}^{(2)}, \quad (45)$$

их зависимости от номера выборки (статистического эксперимента)  $m$  показаны на Рисунке 2. Как видно, это случайные числа с нормальным распределением вероятностей,  $L_1 < L_2$ , соответствующие гистограммы приведены на Рисунке 3, пунктиром показана гауссовская аппроксимация.

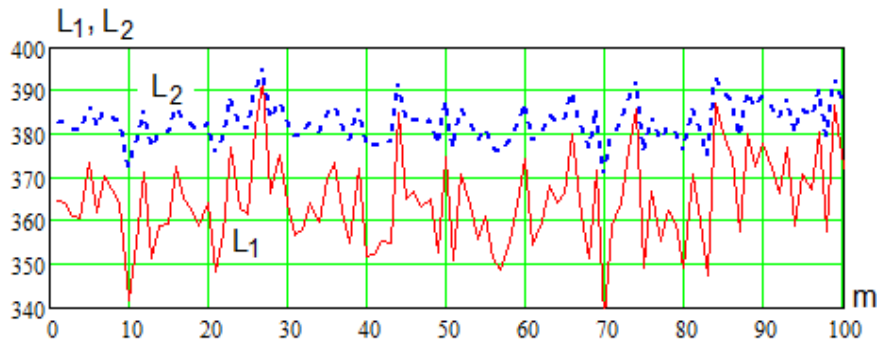


Рисунок 2 – Решающие статистики  
 Figure 2 – Decisive statistics

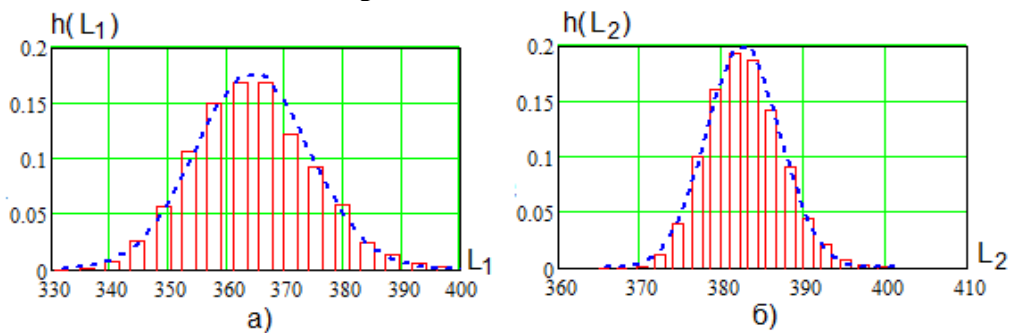


Рисунок 3 – Гистограммы решающих статистик  
 Figure 3 – Decision statistic histograms



В результате статистической обработки реализаций решающих статистик определены их средние значения  $L_{cp1} = 364,5$ ,  $L_{cp2} = 382,5$ , дисперсии  $\sigma_1^2 = 230,7$ ,  $\sigma_2^2 = 165,8$ , среднеквадратические отклонения  $\sigma_1 = 15,2$ ,  $\sigma_2 = 12,9$ . Энтропийные оценки (39), (40) приводят к неравенствам  $\sigma_1^2 < 1448$ ,  $\sigma_2^2 < 1502$  и  $\sigma_1 < 38$ ,  $\sigma_2 < 38,7$ , то есть оказываются приближенными верхними границами. Учет ковариационных составляющих  $C(L_k)$  позволяет уточнить оценку, но требует выборок большого объема, коэффициент корреляции решающих статистик  $L_1$  и  $L_2$  на Рисунке 2 равен 0,934.

Реализация разности решающих статистик  $\Delta L = L_2 - L_1$ , по которой принимается решение о классификации, представлена на Рисунке 4а (пунктиром показаны пороговый уровень  $D$  для вероятностей  $P_0$ , равных 0,99 и 0,999), а ее гистограмма – на Рисунке 4б, пунктиром показана гауссовская аппроксимация.

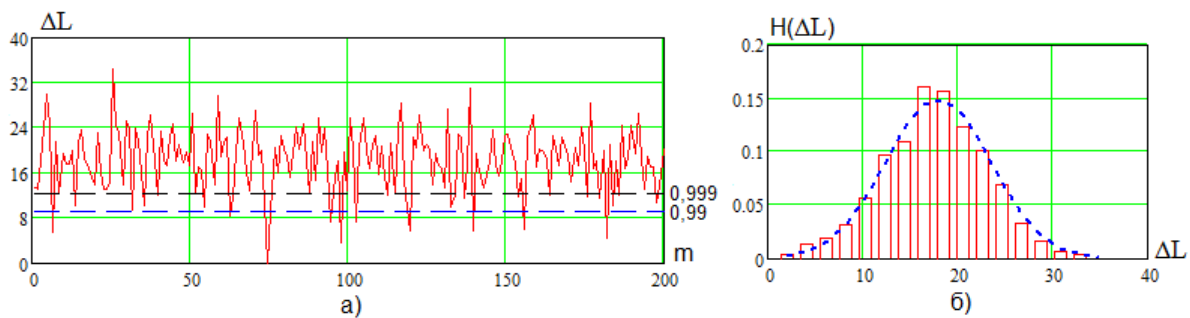


Рисунок 4 – Разность решающих статистик  
 Figure 4 – Difference of decision statistics

Средняя величина  $\Delta L$  равна 18, дисперсия  $\sigma^2(\Delta L) = 32,7$ , среднеквадратическое отклонение  $\sigma_{\Delta L} = 5,72$ , оценки (42), (43) дают те же значения, то есть оказываются достаточно точными. Для выборок из класса  $G_2$  получаются аналогичные результаты.

Согласно (25), (26) для решающих статистик можно записать

$$L_{nm} = (N-1) \cdot \Delta H_{nm}, \quad (46)$$

где  $n=1$  или 2 — номер класса, к которому относится выборка с распределением вероятностей  $Q(i, j)$ ,  $m=1$  или 2 — номер класса, для которого вычисляется решающая статистика,

$$\Delta H_{nm} = H_{2nm}(Q, P^m) - H_{1nm}(Q, P^m). \quad (47)$$

В таблице 1 представлены результаты расчета энтропийных характеристик.

Таблица 1 – Результаты расчета энтропийных характеристик  
 Table 1 – Results of calculating entropy characteristics

m n	$H_{2nm}$		$H_{1nm}$		$\Delta H_{nm}$	
	1	2	1	2	1	2
1	7,371	7,736	3,686	3,868	3,686	3,868
2	9,159	8,534	4,58	4,267	4,58	4,267

Пример реализации последовательности решающих статистик в координатах  $L_1$ ,  $L_2$  в рамках вальдовской процедуры принятия решений, в которой выборка последовательно накапливается по  $N_0=5$  отсчетов, объем выборки равен  $N = n \cdot N_0$  ( $n$  – целое число),  $M=64$ , представлены на Рисунке 5, пунктиром показаны пороговые уровни со значениями  $P_0$ . Точками отмечены результаты статистического имитационного моделирования, прямой линией отображаются энтропийные оценки (46).

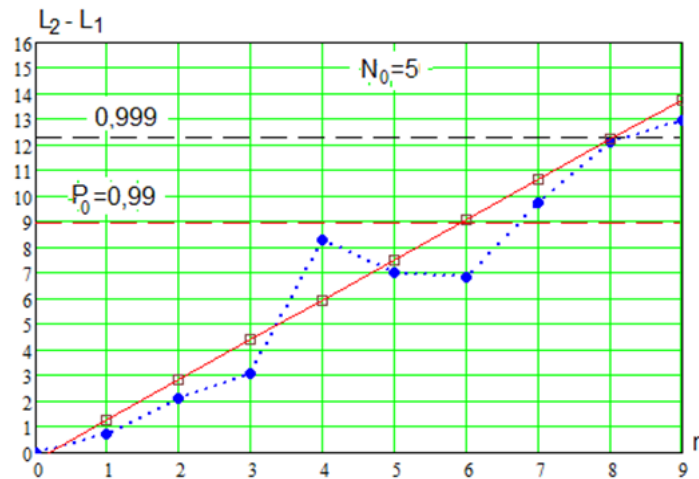


Рисунок 5 – Процесс принятия решения  
 Figure 5 – Decision making process

Как видно, энтропийные оценки решающих статистик и их разностей согласуется с результатами статистического имитационного моделирования. По ним можно оценивать продолжительность процедуры классификации при произвольных вероятностных моделях классов.

### Заключение (Conclusion)

В ходе исследования рассмотрен оптимальный алгоритм классификации с заданной достоверностью случайных сигналов на основе их марковских моделей. Получены энтропийные оценки решающих статистик и их разностей, по которым принимаются решения о принадлежности наблюдаемой выборки отсчетов соответствующему классу. Показано, что они пропорциональны объему выборки отсчетов, собственным энтропиям и взаимным информационным дивергенциям выборки и марковской модели класса. Полученные оценки позволяют анализировать свойства моделей классов и оценивать продолжительность процедуры классификации, которая тем больше, чем меньше взаимная неопределенность сравниваемых классов.

### ЛИТЕРАТУРА

1. Кудряшов Б.Д. *Теория информации*. 2009;19-20.
2. Фукунага К. *Введение в статистическую теорию распознавания образов*. 1979;57-67.
3. Дж. Вэн Райзин *Классификация и кластер*. 1980;7-18.

4. Morris H. DeGroot. *Optimal Statistical Decisions*. Willey Classics Library Edition Published. 200; 230-255.
5. Казаков В.А. *Введение в теорию марковских процессов и некоторые радиотехнические задачи*. 1973;8-12.
6. Тихонов В.И., Миронов М.А. *Марковские процессы*. 1977;69-75.
7. Калинин М.Ю. Особенности разработки программы классификации информационных сигналов на основе марковской модели. *Охрана, безопасность, связь*. 2018;2(3):48-57.
8. Чисар И., Кёрнер Я. *Теория информации*. 1985;35-45.
9. Кульбак С. *Теория информации и статистика*. 1967;95-98.
10. Кендалл М. *Теория распределений*. 1966;477-481.
11. Кендалл М. *Статистические выводы и связи*. 1973;13-19.

### REFERENCES

1. Kudryashov B.D. *Teoriya informatsii*. 2009;19-20.
2. Fukunaga K. *Vvedenie v statisticheskuyu teoriyu raspoznavaniya obrazov*. 1979;57-67.
3. J. Ven Raizin. *Klassifikatsiya i klaster*. 1980;7-18.
4. Morris H. DeGroot. *Optimal Statistical Decisions*. Willey Classics Library Edition Published. 200; 230-255.
5. Kazakov V.A. *Vvedenie v teoriyu markovskih protsessov i nekotorye radiotekhnicheskie zadachi*. 1973;8-12.
6. Tikhonov V.I., Mironov M.A. *Markovskie protsessy*. 1977;69-75.
7. Kalinin M.Y. *Osobennosti razrabotki programmy klassifikatsii informatsionnyh signalov na osnove markovskoj modeli*. 2018;2(3):48-57.
8. Chisar I., Kerner Y. *Teoriya informatsii*. 1985;35-45.
9. Kulbak S. *Teoriya informatsii i statistika*. 1967;95-98.
10. Kendall M. *Teoriya raspredelenij*. 1966;477-481.
11. Kendall M. *Statisticheskie vyvody i svyazi*. 1973;13-19.

### ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

**Калинин Максим Юрьевич**, генеральный директор ООО «Импульс-сервис», Москва, Российская Федерация.  
e-mail: [maks@oxrana.org](mailto:maks@oxrana.org)

**Maxim Y. Kalinin**, general director «Impulse-Service», Moscow, Russian Federation

**Чопоров Олег Николаевич**, д-р техн. наук, профессор кафедры систем информационной безопасности Воронежского государственного технического университета, Воронеж, Российская Федерация  
e-mail: [choporov\\_oleg@mail.ru](mailto:choporov_oleg@mail.ru)  
ORCID: [0000-0002-3176-499X](https://orcid.org/0000-0002-3176-499X)

**Oleg N. Choporov**, Dr. Tech. Sciences, Professor of the Department of Information Security Systems, Voronezh State Technical, Voronezh, Russia