



<http://dx.doi.org/10.35596/1729-7648-2022-20-8-42-50>

Оригинальная статья  
Original paper

УДК 519.2

## ГИСТОГРАММНЫЙ ФИЛЬТР С НАСТРОЙКОЙ ПАРАМЕТРА СГЛАЖИВАНИЯ

А. В. ОВСЯННИКОВ<sup>1</sup>, В. М. КОЗЕЛ<sup>2</sup>

<sup>1</sup>Белорусский государственный университет (г. Минск, Республика Беларусь)

<sup>2</sup>Белорусский государственный университет информатики и радиоэлектроники  
(г. Минск, Республика Беларусь)

Поступила в редакцию 18.04.2022

© Белорусский государственный университет информатики и радиоэлектроники, 2022  
Belarusian State University of Informatics and Radioelectronics, 2022

**Аннотация.** В статье рассматривается гистограммный фильтр с настройкой параметра сглаживания. Гистограммный фильтр может быть эффективно применен в задачах идентификации (распознавания) законов распределения для малых объемов данных. Параметр сглаживания определяется с учетом имеющейся в наличии априорной информации относительно предполагаемого закона распределения. Установлено соотношение между математическими ожиданиями критерия согласия хи-квадрат стандартной гистограммной оценки и с использованием гистограммного фильтра. Такое соотношение определяется коэффициентом сглаживания. Численное значение коэффициента сглаживания зависит от параметров: объема данных, количества интервалов группирования данных, параметров формы закона распределения. Проведен анализ целесообразности применения гистограммного фильтра с учетом соотношения указанных выше параметров. Зависимость коэффициента сглаживания от этих параметров позволяет определить взаимосвязь между количеством интервалов группирования данных и их объемом. Гистограммный фильтр является простым для реализации инструментом, который легко может быть встроен в любой открытый алгоритм идентификации (распознавания) закона распределения.

**Ключевые слова:** гистограммный фильтр, идентификация, коэффициент сглаживания, объем данных, интервал группирования, плотность распределения вероятности.

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

**Для цитирования.** Овсянников, А. В. Гистограммный фильтр с настройкой параметра сглаживания / А. В. Овсянников, В. М. Козел // Доклады БГУИР. 2022. Т. 20, № 8. С. 42–50. <http://dx.doi.org/10.35596/1729-7648-2022-20-8-42-50>.

## HISTOGRAM FILTER WITH SMOOTHING PARAMETER SETTING

ANDREI V. AUSIANNIKAU<sup>1</sup>, VICTOR M. KOZEL<sup>2</sup>

<sup>1</sup>Belarusian State University (Minsk, Republic of Belarus)

<sup>2</sup>Belarusian State University of Informatics and Radioelectronics (Minsk, Republic of Belarus)

Submitted 18.04.2022

**Abstract.** A histogram filter with smoothing parameter settings is discussed in the article. The histogram filter can be effectively applied in the problems of identification (recognition) of distribution laws for small amounts of data. The smoothing parameter is determined taking into account the available a priori information regarding the proposed distribution law. The relationship between the mathematical expectations of the chi-square fit criterion of the standard estimation histogram and the use of the histogram filter has been determined. This ratio is determined by the smoothing factor. The numerical value of the smoothing coefficient depends on the following parameters: the amount of data, the number of grouping intervals, and the shape parameters of the distribution law. The paper analyzes the feasibility of using a histogram filter, depending on the ratio of the above parameters. The dependence of the smoothing coefficient on the specified parameters allows one to determine the relationship between the number of data grouping intervals and their volume. The histogram filter is an easy-to-implement tool that can be easily integrated into any open distribution law identification (recognition) algorithm.

**Keywords:** histogram filter, identification, smoothing coefficient, data volume, grouping interval, probability density distribution.

**Conflict of interests.** The authors declare no conflict of interests.

**For citation.** Ausiannikau A. V., Kozel V. M. (2022) Histogram Filter with Smoothing Parameter Setting. *Doklady BGUIR*. 20 (8), 42–50. <http://dx.doi.org/10.35596/1729-7648-2022-20-8-42-50> (in Russian).

### Введение

Проблематика гистограммных оценок плотности распределения вероятности (ППВ) хорошо известна: отсутствие единых взглядов на определение числа интервалов группирования данных (Р 50.1.033–2001 «Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть 1. Критерии типа хи-квадрат») и сильная изрезанность гистограммы при относительно малом числе наблюдений [1, 2]. Особую важность и актуальность точные гистограммные оценки закона распределения приобретают в случае требований его быстрой идентификации (распознавания), возможной высокой стоимости измерительного эксперимента или процесса получения данных. Требование быстрой идентификации (распознавания) закона распределения в данном случае можно считать эквивалентным малому объему данных.

Устранение проблем изрезанности гистограммы заключается в применении гистограммных фильтров (ГФ), например, усредняющего, медианного, гауссовского и др. [1, 3–5]. В то же время их использование эмпирически интуитивно и исходит в основном из практических соображений. В статье предлагается теоретически обоснованная методика реализации ГФ, работающего на небольшом количестве данных, устраняющего изрезанность гистограммы, дающего «правильную» идентификацию закона распределения, ослабляющего зависимость «правильной» идентификации от числа интервалов группирования данных. Также развиты идеи, предложенные в [6]. Прежде всего, рекомендуется отказаться от единичной функции включения данных в интервал группирования: данные могут находиться вблизи границ интервала и при изменении числа интервалов оказаться в соседнем интервале; для относительно небольшого количества данных устранение эффекта изрезанности гистограммы может быть осуществлено сглаживанием данных на нескольких соседних интервалах. В этом случае целесообразно заменить единичную функцию включения взвешенной функцией, учитывающей возможный вес «ошибочно» попавших в соседние интервалы данных:  $u_j = \alpha_j v_{j-1} + k_j v_j + \beta_j v_{j+1}$ ,  $\alpha_j + k_j + \beta_j = 1$ , где  $v_j$  – число данных, попавших в  $j$ -й интервал группирования;  $\{\alpha_j, k_j, \beta_j\}$  – весовые коэффициенты интервалов (параметры сглаживания). В простейшем случае весовые коэффициенты являются постоянными величинами и могут быть выражены через один коэффициент  $k$  – параметр сглаживания. Введение весовых

коэффициентов для малых объемов данных позволяет перегруппировать эти данные так, чтобы обеспечить меньшую изрезанность гистограммы, увеличив при этом ее сглаженность и быструю идентификацию.

Вычисление параметра сглаживания, очевидно, требует некоторой априорной информации об идентифицируемой ПРВ. Принято, что такая идентификация проводится с помощью критерия согласия хи-квадрат, использование которого основано на предположении о возможном виде идентифицируемой ПРВ. Таким образом, априорная информация является естественным и необходимым элементом построения ГФ.

Цель исследований состояла в реализации гистограммного фильтра с настройкой параметра сглаживания на основе априорной информации об идентифицируемой ПРВ.

### Общая методика построения гистограммного фильтра с настройкой параметра сглаживания

Пусть имеется выборка случайных данных  $\{x_i\}$ ;  $i = \overline{1, n}$  и определено разбиение числовой прямой на  $m$  непересекающихся и примыкающих друг к другу интервалов  $A_j$ ;  $j = \overline{1, m}$  равной длины  $\Delta_x = X_{j+1} - X_j = (x_{\max} - x_{\min}) / m$ ;  $X_{m+1} = x_{\max} = \max_i x_i$ ;  $X_1 = x_{\min} = \min_i x_i$ , где  $X_j$  – границы интервалов;  $R = x_{\max} - x_{\min} = m\Delta_x$  – размах диапазона данных. Заменим обычную индикаторную функцию, используемую при стандартном способе построения гистограммы, весовой ступенчатой функцией  $\mu_j(x_i)$ ,  $0 \leq \mu_j \leq 1$ ,  $\Delta_\mu \geq \Delta_x$ , которая будет характеризовать принадлежность данных интервалу группирования  $A_j$ :

$$\begin{cases} \mu_j(x) = \{k \text{ для } A_j; \alpha = (1-k)/2 \text{ для } A_{j-1} \text{ и } A_{j+1}\}, & j = \overline{2, m-1}; \\ \mu_j(x) = \{(1+k)/2 \text{ для } A_j; \alpha \text{ для } A_{(j-m)(m-3)/(m-1)+(m-1)}\}, & j = 1, m. \end{cases} \quad (1)$$

Такой выбор весовых коэффициентов обеспечивает выполнение общих условий нормировки:

$$1) \sum_{t=j-1}^{j+1} \mu_t = 1, \quad t = \overline{2, m-1}; \quad 2) \sum_j \sum_i \mu_j(x_i) = n. \quad (2)$$

Используя в качестве критерия оценки коэффициента сглаживания критерий хи-квадрат и заменив число  $v_j$  в критерии  $\chi^2(v)$  на число  $u_j = \alpha v_{j-1} + kv_j + \alpha v_{j+1}$  для  $j = \overline{2, m-1}$  и  $u_j = (1-\alpha)v_j + \alpha v_{(j-m)(m-3)/(m-1)+(m-1)}$  для  $j = 1, m$ , получим

$$\chi_{\text{ГФ}}^2(u) = \sum_{j=1}^m [u_j - np_j]^2 / np_j \rightarrow \min_k. \quad (3)$$

Решение оптимизационной задачи (3) приводит к выражению для коэффициента сглаживания по выборке данных

$$k_{\text{выб}} = 1 + 2 \left[ \sum_{j=1}^m U_j^2 / np_j \right]^{-1} \sum_{j=1}^m (v_j - np_j) U_j / np_j = 1 + 2 \left[ \sum_{j=1}^m U_j^2 / np_j \right]^{-1} \sum_{j=1}^m v_j U_j / np_j, \quad (4)$$

где  $U_j = v_{j-1} - 2v_j + v_{j+1}$ ,  $U_j = -v_j + v_{(j-m)(m-3)/(m-1)+(m-1)}$  – конечная разность второго порядка соответственно для индексов  $j = \overline{2, m-1}$  и  $j = 1, m$ ;  $\sum_{j=1}^m U_j = 0$ ;  $p_j$  – гипотетическая вероятность.

Статистическая конкретизация формулы (4) приводит к соотношению

$$k = 1 - \frac{4(m-1)}{6(m-1) + n\Delta_x^4 A(f, f^*)} = 1 - \frac{1}{1,5 + 0,25n\Delta_x^4 (m-1)^{-1} A(f, f^*)}, \quad (5)$$

где  $A(f, f^*) = \int_R (f^{**} / f)^2 f dx$  – коэффициент формы ПРВ, эквивалентный информации Фишера второго порядка [7];  $f$  – гипотетическая ПРВ;  $f^* = \lim_{j \rightarrow \infty, \Delta_x \rightarrow 0} [u_j / n\Delta_x]$  – эквивалент идентифицируемой ПРВ;  $\int_R f dx = \gamma$  – доверительная вероятность.

Проведем упрощенное обоснование формулы (5), для чего последовательно рассмотрим компоненты, входящие в (4). Совокупность статистическо-экспериментального метода, инженерного подхода и практических представлений приводит к следующим выражениям:

$$\begin{aligned} \sum_{j=1}^m U_j / np_j &= \sum_{j=2}^{m-1} (f_{j-1}^* - 2f_j^* + f_{j+1}^*) / f_j + (-f_1^* + f_2^*) / f_1 + (-f_m^* + f_{m-1}^*) / f_m; \\ \lim_{\substack{m \rightarrow \infty \\ \Delta_x \rightarrow 0}} \sum_{j=1}^m U_j / np_j &= (R/m) \int_{x_{\min} + \Delta_x}^{x_{\max} - \Delta_x} f^{**} / f dx + \Delta_x (f_1^* / f_1 - f_m^* / f_m); \\ \mathbf{M} \left( \sum_{j=1}^m \frac{v_j U_j}{np_j} \right) &= -2(n+m-1) + \mathbf{M} \left( \sum_{j=2}^{m-1} \frac{(v_{j-1} + v_{j+1})v_j}{np_j} + \frac{v_1 U_1}{np_1} + \frac{v_m U_m}{np_m} \right) = -2(m-1) - 2\mathbf{M}(v_1 + v_m); \\ \lim_{\substack{m \rightarrow \infty \\ \Delta_x \rightarrow 0}} \mathbf{M} \left( \sum_{j=1}^m \frac{v_j U_j}{np_j} \right) &= \lim_{\substack{m \rightarrow \infty \\ \Delta_x \rightarrow 0}} \mathbf{M} \left( -2(m-1) - 2 \int_{\Delta_x} f^* dx \right) = -2(m-1); \end{aligned} \quad (6)$$

$$\begin{aligned} \lim_{\substack{m \rightarrow \infty \\ \Delta_x \rightarrow 0}} \mathbf{M} \left( \sum_{j=1}^m \frac{U_j^2}{np_j} \right) &= \lim_{\substack{m \rightarrow \infty \\ \Delta_x \rightarrow 0}} \mathbf{M} \left( \sum_{j=2}^{m-1} \frac{v_{j-1}^2 + 4v_j^2 + v_{j+1}^2}{np_j} \right) - 2 \lim_{\substack{m \rightarrow \infty \\ \Delta_x \rightarrow 0}} \mathbf{M} \left( \sum_{j=2}^{m-1} \frac{2v_{j-1}v_j - v_{j-1}v_{j+1} + 2v_jv_{j+1}}{np_j} \right) + \\ &+ \lim_{\substack{m \rightarrow \infty \\ \Delta_x \rightarrow 0}} \mathbf{M} \left( \frac{(-v_1 + v_2)^2}{np_1} + \frac{(-v_m + v_{m-1})^2}{np_m} \right) \approx 6(n+m-1) - 6n + n\Delta_x^4 \int_R \left( \frac{f^{**}}{f} \right)^2 f dx = \\ &= 6(m-1) + n\Delta_x^4 A(f, f^*). \end{aligned} \quad (7)$$

Далее, подставляя (6) и (7) в (4), получим непосредственно выражение (5). Формула (5) позволяет сделать ряд важных выводов.

Во-первых, при неограниченно возрастающем числе данных  $n \rightarrow \infty$  очевидно, что коэффициент сглаживания должен стремиться к единице, что и следует из формулы (5). В этом случае целесообразность применения ГФ исчезает. При значении компоненты знаменателя  $\delta = 0,25n\Delta_x^4(m-1)^{-1}A(f, f^*)$  меньше единицы или при  $\delta \rightarrow 0$  коэффициент сглаживания стремится к  $1/3$ . Такое значение коэффициента сглаживания отвечает случаю сильной изрезанности гистограммы; возможно, вследствие неправильно выбранного (относительно большого) значения количества интервалов при относительно небольшом количестве данных. Гистограммный фильтр в этом случае преобразуется в обычный усредняющий фильтр. Таким образом, диапазон изменения значений коэффициента сглаживания  $1/3 \leq k \leq 1$ .

Во-вторых, подставляя значение коэффициента сглаживания (5) в формулу для критерия согласия хи-квадрат, получаем  $\chi_{\text{гф}}^2(u) = \chi^2(v) - \left( \sum_{j=1}^m U_j^2 / np_j \right)^{-1} \left( \sum_{j=1}^m v_j U_j / np_j \right)^2$ . Отсюда следует соотношение между средними критерия хи-квадрат:  $\mathbf{M}(\chi_{\text{гф}}^2) = k\mathbf{M}(\chi^2)$ ,  $\mathbf{M}(\chi^2) = m-1$ .

Таким образом, применение ГФ позволяет уменьшить значение стандартного критерия согласия в  $k$  раз. Соотношение входящих в коэффициент параметров характеризует целесообразность применения и эффективность идентификации ГФ. Так, при небольших значениях компоненты знаменателя  $\delta < 1$  значение критерия хи-квадрат после применения фильтра практически уменьшается в три раза, в противном случае при  $n \rightarrow \infty$  значение критерия хи-квадрат стремится к стандартному ( $k \rightarrow 1$ ,  $\mathbf{M}(\chi_{\text{гф}}^2) \rightarrow (m-1)$ ), и применение ГФ нецелесообразно. Следовательно, эффективность ГФ можно оценивать величиной, обратной значению коэффициента сглаживания:  $\mathfrak{E}_{\text{гф}} = k^{-1}$ .

Показатель эффективности может быть использован в целях нахождения оптимального соотношения между числом данных и числом интервалов их группирования. Зафиксировав некоторое желаемое значение эффективности  $\mathfrak{E}_{\text{гф}}^0$ , на основании (5) получаем

$$\frac{n}{m^4(m-1)} = \frac{2K}{R^4 A}, \quad K = \frac{3 - \mathfrak{E}_{\text{гф}}}{\mathfrak{E}_{\text{гф}} - 1} = \frac{3k - 1}{1 - k}. \quad (8)$$

Уравнение (8) нелинейное, требующее численных методов решения в общем случае. Из формулы следует, что число интервалов группирования данных зависит не только от числа данных, но и от коэффициента формы ПРВ  $A(f, f^*)$ , и от диапазона данных  $R$ . Для случая  $m \gg 1$  возможно приближенное аналитическое решение нелинейного уравнения (8):  $m = \sqrt[3]{nR^4 A(f, f^*) / 2K}$ . Последняя формула близка к выражениям вида  $m \sim n^{0,2}$ , приведенным в [1, 8] с коэффициентом пропорциональности, зависящим от параметров ПРВ и априорных установок ГФ по его эффективности  $\Theta_{\text{ГФ}}^0$ .

В-третьих, перераспределение части данных между соседними интервалами не только уменьшает изрезанность гистограммы, но и способствует ослаблению требований к выбору числа интервалов группирования. Фиксируя некоторую нижнюю границу значения эффективности ГФ ( $\Theta_{\text{ГФ}}^0$ ), можно определить значение числа интервалов группирования по формуле (8) из условия  $m^4(m-1) \geq 0,5nR^4 A(f)(K)^{-1}$ .

В-четвертых, предположив высокую апостериорную точность оценки ПРВ, плотность  $f^*$  в коэффициенте формы формально можно заменить гипотетической и, следовательно, модифицированный коэффициент формы примет вид  $A(f) = \int_R (f'' / f)^2 f dx$ . В этом случае формулы (5) и (8) становятся полностью определенными. Следует заметить, что вычисление этого коэффициента требует существования второй производной ПРВ. Однако, если такой производной нет, нужно воспользоваться инженерными соображениями практической реализации. В частности, для равномерной ПРВ можно принять  $f'' = 0$  и, следовательно, коэффициент формы  $A(f) = 0$ . Тогда численное значение коэффициента сглаживания будет  $1/3$ , и ГФ преобразуется в обычный усредняющий фильтр, что в случае идентифицируемой равномерной ПРВ вполне очевидно.

### Рекомендации по реализации гистограммного фильтра

Полученные теоретические результаты показывают целесообразность применения ГФ с целью эффективной и быстрой (на малых объемах данных) идентификации изменяющихся законов распределения в описательной статистике при обработке гистограмм изображений. Программная реализация ГФ легко встраивается в существующие открытые алгоритмы построения гистограмм, например, в функции `hist`, `histfit` платформы MATLAB.

Структура алгоритма идентификации (распознавания) ПРВ следующая.

1. Получение выборки данных объемом  $n$ , определение размаха выборки  $R$ .
2. На основании предположений об идентифицируемой ПРВ вычисление коэффициента формы  $A(f)$ .
3. На основании выбранного числа интервалов группирования данных, размаха выборки, объема данных, коэффициента формы вычисляется значение коэффициента сглаживания по формуле (5).
4. Применение ГФ:

$$\begin{cases} u_j = \alpha v_{j-1} + k v_j + \alpha v_{j+1}, & j = \overline{2, m-1}; \\ u_j = (1 - \alpha) v_j + \alpha v_{(j-m)(m-3)/(m-1)+(m-1)}, & j = 1, m; \\ \alpha = (1 - k) / 2. \end{cases} \quad (9)$$

5. Вычисление критерия согласия хи-квадрат. На основании заданного уровня значимости принятие решения об идентификации.

Процедуру идентификации ПРВ можно сделать многоканальной, где каждый канал будет ориентирован на определенный заранее возможный вид ПРВ. Принятие решения об идентификации в этом случае может быть реализовано различными методами, например, простым или взвешенным голосованием.

### Моделирование гистограммного фильтра

На рис. 1 ( $n = 100$ ,  $m = 9$ , количество серий экспериментов 100) приведены примеры работы ГФ для ПРВ: нормальной (рис. 1,  $a$ ,  $k_{\text{выб}} = 0,64$ ,  $\Theta_{\text{выб}} = 1,56$ ), логистической (рис. 1,  $b$ ,  $k_{\text{выб}} = 0,74$ ,



$\mathfrak{E}_{\text{выб}} = 1,35$ ), экспоненциальной (рис. 1,  $c$ ,  $k_{\text{выб}} = 0,53$ ,  $\mathfrak{E}_{\text{выб}} = 1,89$ ), равномерной (рис. 1,  $d$ ,  $k_{\text{выб}} = 0,36$ ,  $\mathfrak{E}_{\text{выб}} = 2,78$ ). Верхняя часть на рис. 1 соответствует обычной гистограмме, нижняя – результат обработки ГФ. Во всех приведенных на рисунке случаях выполняется соотношение  $\chi_{\text{ГФ}}^2 < \chi_{\text{кр}}^2 \leq \chi^2$ , где  $\chi_{\text{кр}}^2$  – критическое значение критерия согласия при заданном уровне значимости (0,05). Результаты моделирования наглядно подтверждают идею применения ГФ. Эффективность использования гистограммного фильтра на отмеченных плотностях указывает на существенное его превосходство перед стандартной гистограммной оценкой.

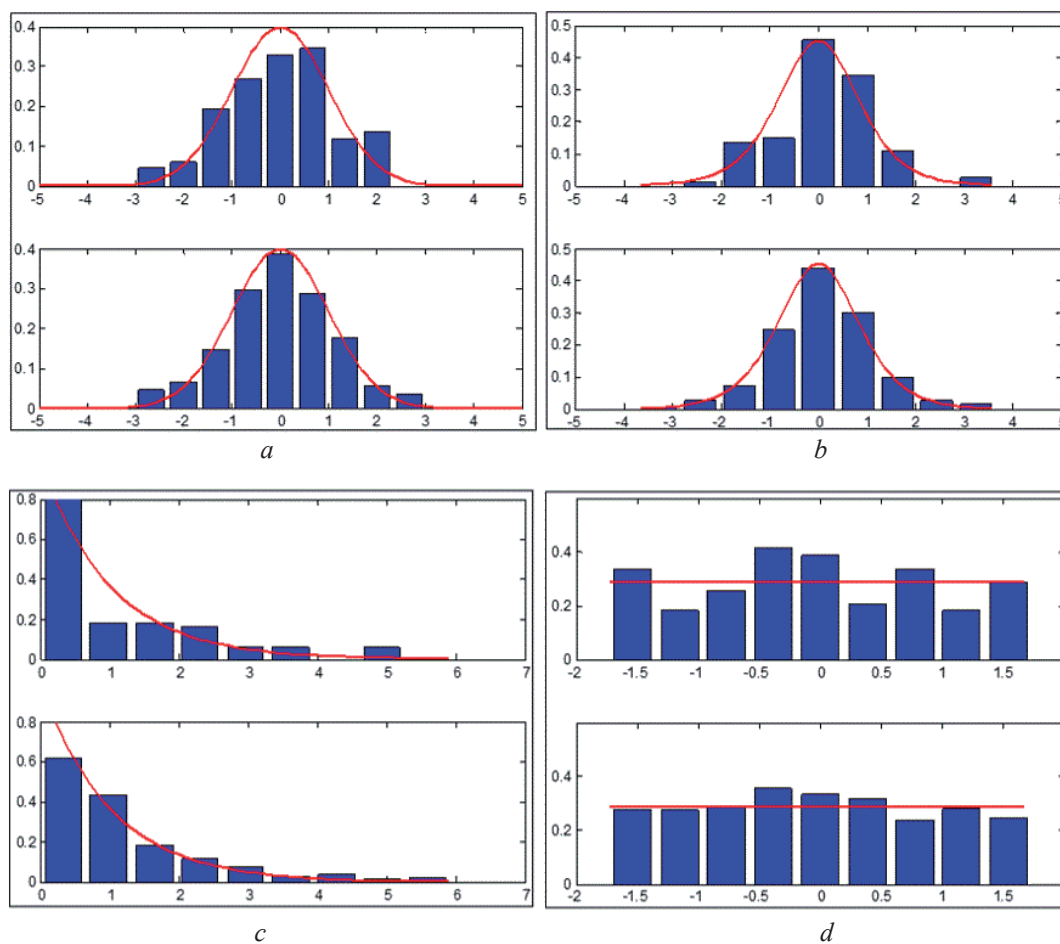


Рис. 1. Результаты работы гистограммного фильтра  
Fig. 1. Histogram filter results

В табл. 1 (количество серий экспериментов 100) содержатся некоторые результаты моделирования работы ГФ с учетом формулы (4) в сопоставлении с теоретическими результатами, полученными на основе (5) для трех ПРВ: нормальной, логистической, лапласовской.

Таблица 1. Коэффициенты сглаживания и эффективность гистограммного фильтра,  $n = 100$   
Table 1. Smoothing coefficients and histogram filter efficiency,  $n = 100$

$m$	Плотность распределения вероятности								
	нормальная, $A(f) = 1,73$			логистическая, $A(f) = 2,14$			лапласовская, $A(f) = 0,99$		
	$k_{\text{выб}}$	$k$	$\mathfrak{E}_{\text{ГФ}}$	$k_{\text{выб}}$	$k$	$\mathfrak{E}_{\text{ГФ}}$	$k_{\text{выб}}$	$k$	$\mathfrak{E}_{\text{ГФ}}$
5	0,80	0,96	1,04	0,96	0,98	1,02	0,98	0,99	1,01
7	0,81	0,81	1,23	0,90	0,92	1,09	0,94	0,97	1,03
9	0,66	0,61	1,64	0,75	0,77	1,30	0,86	0,91	1,10

На рис. 2 приведены зависимости коэффициента сглаживания, вычисленного по формуле (5), от количества интервалов группирования  $n$  для двух ПРВ: нормальной (кривые 1, 2) и лапласовской (кривые 3, 4). Численные (табл. 1) и графические результаты (рис. 2) позволяют сделать вывод о том, что значение коэффициента сглаживания нелинейно уменьшается с уменьшением объема данных. Это объясняется тем, что при уменьшающемся объеме данных увеличивается изрезанность обычной гистограммной оценки ПРВ и, следовательно, требуется ее большая сглаженность, стремящаяся к равномерному (усредняющему) сглаживанию ( $k \rightarrow 1/3$ ) при одном и том же числе интервалов группирования.

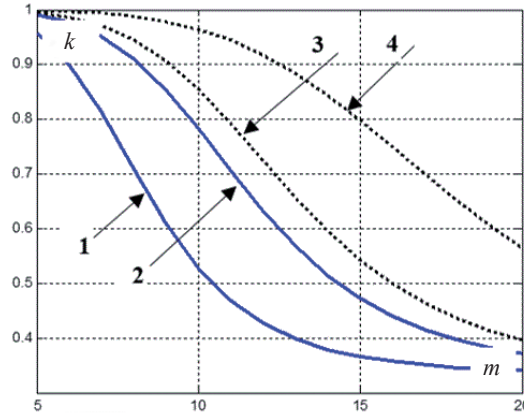


Рис. 2. Коэффициенты сглаживания гистограммного фильтра: 1, 3 –  $n = 100$ ; 2, 4 –  $n = 500$   
Fig. 2. Histogram filter smoothing coefficients: 1, 3 –  $n = 100$ ; 2, 4 –  $n = 500$

На рис. 3 (количество серий экспериментов 100) на основе формулы (8) (кривые 1.1–1.3) приведены зависимости между объемом данных и количеством интервалов их группирования для двух плотностей – гауссовской (рис. 3, *a*) и логистической (рис. 3, *b*) – для различных значений коэффициента эффективности:  $\mathfrak{E}_{\text{гф}} = 1,5$  ( $k = 0,6$ ) – кривая 1.1;  $\mathfrak{E}_{\text{гф}} = 1,3$  ( $k = 0,77$ ) – кривая 1.2;  $\mathfrak{E}_{\text{гф}} = 1,1$  ( $k = 0,91$ ) – кривая 1.3. На рис. 3 для сопоставления приведены стандартно используемые формулы Старджеса  $m = 1 + \log_2 n$  (кривая 2) и формула  $m = C(E_x)n^{0,4}$  (кривая 3), приведенная в [1, 8], где  $C(E_x) = (E_x + 4,5)/6$ ;  $E_x$  – коэффициент эксцесса ПРВ.

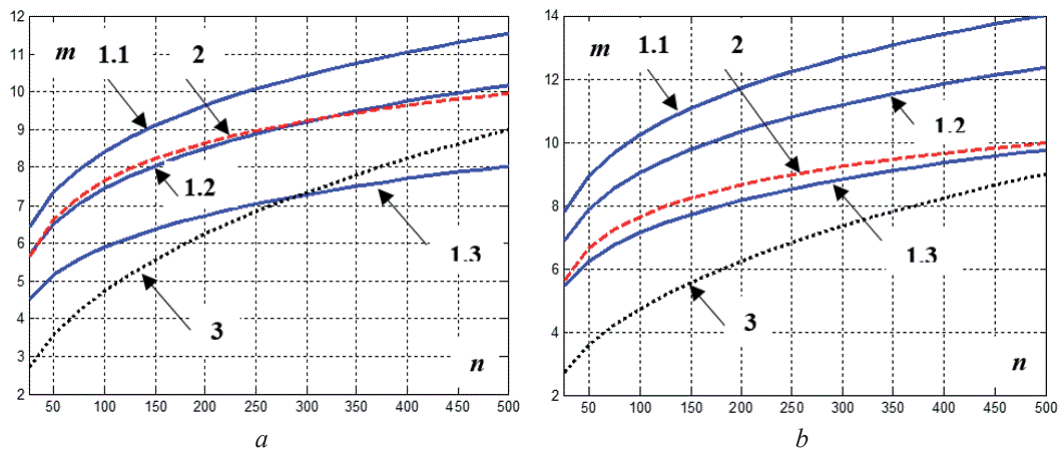


Рис. 3. Зависимость количества интервалов группирования данных от объема данных  
Fig. 3. Dependence of the number of grouping intervals of data on the amount of data

### Закключение

1. Рассмотренный гистограммный фильтр (9) с настройкой параметра сглаживания может быть эффективно применен в задачах идентификации (распознавания) плотности распределения вероятности для малых объемов данных с учетом имеющейся в наличии априорной информации о предполагаемой плотности распределения вероятности.

2. Установлено соотношение между математическими ожиданиями согласия критерия хи-квадрат при стандартном подходе построения гистограммной оценки и с использованием гистограммного фильтра. Такое соотношение определяется коэффициентом сглаживания, определяемым по формуле (5). Численное значение коэффициента сглаживания зависит от следующих параметров: объема данных, количества интервалов группирования данных, параметров формы плотности распределения вероятности. Зависимость коэффициента сглаживания от указанных параметров позволяет определить взаимосвязь между количеством интервалов группирования данных и их объемом. Эта зависимость нелинейная, не имеющая аналитического решения. В случае выбранного большого количества интервалов группирования данных может быть получена приближенная формула в явном виде (8).

3. Гистограммный фильтр является простым для реализации инструментом, который легко может быть встроен в любой открытый алгоритм идентификации (распознавания) плотности распределения вероятности.

### Список литературы

1. Орлов, Ю. Н. Оптимальное разбиение гистограммы для оценивания выборочной плотности функции распределения нестационарного временного ряда / Ю. Н. Орлов // Препринты ИПМ имени М. В. Келдыша. 2013. № 14. 26 с. <http://library.keldysh.ru/preprint.asp?id=2013-14>.
2. Chong, Gu. Nonparametric Density Estimation in High-Dimensions / Gu. Chong, Jeon Yongho, Lin Yi // *Statistica Sinica*. 2013. No 23. P. 1131–1153.
3. Devroye, L. Nonparametric Density Estimation: the L1 View / L. Devroye, L. Györfi. New York: John Wiley Sons, 1985.
4. Solomon, C. J. Fundamentals of Digital Image Processing: a Practical Approach with Examples in Matlab / C. J. Solomon, T. P. Breckon. USA: Wiley-Blackwell, 2010. DOI: 10.1002/9780470689776.
5. Gonzalez, R. Digital Image Processing / R. Gonzalez. New York: Pearson, 2018.
6. Овсянников, А. В. Фильтрация гистограммной оценки плотности вероятности на основе нечеткой принадлежности данных интервалу группирования / А. В. Овсянников, В. М. Козел // Доклады БГУИР. 2021. Т. 19, № 4. С. 13–20. <https://doi.org/10.35596/1729-7648-2021-19-4-13-20>.
7. Овсянников, А. В. Статистические неравенства в сверхрегулярных статистических экспериментах теории оценивания / А. В. Овсянников // Весті Нацыянальнай акадэміі навук Беларусі. Сер. фіз.-мат. навук. 2009. № 2. С. 106–110.
8. Новицкий, П. В. Оценка погрешностей результатов наблюдений / П. В. Новицкий, И. А. Зограф.; 2-е изд. перераб. и доп. Л.: Энергоатомиздат, Ленингр. отд-ние, 1991. 304 с.

### References

1. Orlov Yu. N. (2013) Optimal Histogram Partitioning for Estimating the Sample Density of the Distribution Function of an Unsteady Time Series. *Preprinty IPM imeni M. V. Keldysha = Preprints of the Institute of Applied Mathematics name after M. V. Keldysh RAS.* (14), 26. <http://library.keldysh.ru/preprint.asp?id=2013-14>.
2. Chong Gu, Yongho Jeon, Yi Lin (2013) Nonparametric Density Estimation in High-Dimensions. *Statistica Sinica*, 23, 1131–1153.
3. Devroye L., Györfi L. (1985) *Nonparametric Density Estimation: the L1 View*. New York, John Wiley Sons Publ.
4. Solomon C. J., Breckon T. P. (2010) *Fundamentals of Digital Image Processing: a Practical Approach with Examples in Matlab*. USA, Wiley-Blackwell Publ. DOI: 10.1002/9780470689776.
5. Gonzalez R. (2018) *Digital Image Processing*. New York, Pearson Publ.
6. Ausiannikau A. V., Kozel V. M. (2021) Filtration of Histogram Evaluation of Probability Density Based on Fuzzy Data Accessibility to a Grouping Interval. *Doklady BGUIR*. 19(4), 13–20. <https://doi.org/10.35596/1729-7648-2021-19-4-13-20> (in Russian).
7. Ausiannikau A. V. (2009) Statisticheskie Neravenstva v Sverh Regulyarnyh Statisticheskikh Eksperimentah Teorii Otsenivaniya. *Vestsi Natsiyanal'naj Akademii Navuk Belarusi. Ser. Fiz-Mat. Navuk.* (2), 106–110 (in Russian).
8. Novitskiy P. V., Zograf I. A. (1991) *Evaluation of Errors of Measurement Results, 2-e izd. pererab. i dop.* Leningrad, Energoatomizdat Publ., Leningrad. Otdelenie. 304 (in Russian).



### **Вклад авторов**

Овсянников А. В. разработал общую методику построения гистограммного фильтра с настройкой параметра сглаживания.

Козел В. М. выполнил моделирование работы гистограммного фильтра.

### **Authors' contribution**

Ausiannikau A. V. developed a general technique for constructing a histogram filter with adjustment of the smoothing parameter.

Kozel V. M. performed modeling of the histogram filter.

### **Сведения об авторах**

**Овсянников А. В.**, к.т.н., доцент, доцент кафедры информационных технологий Белорусского государственного университета.

**Козел В. М.**, к.т.н., доцент, доцент кафедры информационных радиотехнологий Белорусского государственного университета информатики и радиоэлектроники.

### **Адрес для корреспонденции**

220030, Республика Беларусь,  
г. Минск, просп. Независимости, 4  
Белорусский государственный университет  
Тел. +375 17 209-58-94  
E-mail: andovs@mail.ru  
Овсянников Андрей Витальевич

### **Information about the authors**

**Ausiannikau A. V.**, Cand. of Sci., Assistant Professor, Assistant Professor at the Department of Information Technologies of the Belarusian State University.

**Kozel V. M.**, Cand. of Sci., Assistant Professor, Assistant Professor at the Department of Information Radiotechnologies of the Belarusian State University of Informatics and Radioelectronics.

### **Address for correspondence**

220030, Republic of Belarus,  
Minsk, Nezavisimosti Ave., 4  
Belarusian State University  
Tel. +375 17 209-58-94  
E-mail: andovs@mail.ru  
Ausiannikau Andrei Vital'evich