

## Репрезентативность статистической выборки при количественном анализе изображений

В. Н. Петрушин<sup>\*</sup>, Ю. В. Рудяк<sup>\*\*</sup>, Г. О. Рытиков<sup>\*\*</sup>

*<sup>\*</sup>Институт современного искусства  
121309, Москва, ул. Новозаводская, 27А*

*<sup>\*\*</sup>Московский политехнический университет  
107023, Москва, ул. Б. Семеновская, 38*

*e-mail: petrushinon@mail.ru*

*Аннотация.* Статья посвящена обсуждению концепции понятия «репрезентативность статистической выборки». Предложена статистическая мера репрезентативности и способ количественной оценки степени репрезентативности выборки из генеральной совокупности. Для демонстрации техники применения соответствующих определений и методов рассмотрен пример, связанный с выявлением объема статистически достоверной выборки при анализе СЭМ-изображений поверхности полимерного композита.

*Ключевые слова:* генеральная совокупность, статистическая выборка, плотность вероятности, эмпирическое распределение, гистограмма, достоверность, анализ изображений.

### 1. Введение

При описании и моделировании результатов научных и производственных исследований одним из наиболее известных и широко применяемых является статистический подход. Совокупность количественных методов статистического подхода опирается преимущественно на понятийный аппарат и многочисленные техники применения элементов теории вероятностей и математической статистики, объединяемые понятием «прикладная статистика». Базовыми понятиями прикладной статистики являются генеральная и выборочная совокупности [1–7].

Вследствие классической гносеологической проблемы (невозможность формального определения фундаментальных понятий) единого и общепринятого определения генеральной совокупности и статистической выборки, по всей видимости, не существует. Тем не менее многочисленные исследователи предметной области при употреблении обсуждаемых понятий успешно понимают друг друга на интуитивном уровне. Под генеральной совокупностью обычно понимается все «целевое» множество объектов, относительно которого должны быть получены некоторые выводы или прогнозы по результатам статистического исследования. Выборочной совокупностью называют то множество объектов, над которыми непосредственно

удается осуществить измерения и/или вычисления, на основании которых формулируются те или иные статистические модели, с помощью которых формулируются выводы и делаются прогнозы относительно генеральной совокупности.

При этом неявно предполагается, что модели, успешно описывающие выборку, будут эффективны и для описания генеральной совокупности. Вся критика методов прикладной статистики вообще и параметрической статистики в частности основана на невозможности доказательства или опровержения этого предположения. Для того, чтобы снять обсуждаемую проблему, при статистическом анализе данных разными способами вводится понятие «оценка репрезентативности выборки», которое постулирует, что при формировании количественных статистических моделей используется выборка, «похожая» на генеральную совокупность в такой степени, что результаты прогнозов для выборки и для генеральной совокупности должны совпадать. Очевидно, что проверить это утверждение также невозможно в том случае, если вся генеральная совокупность не известна. Если же вся генеральная совокупность известна, то, вообще говоря, необходимости в применении статистических методов для ее описания зачастую нет.

Вместе с тем прогнозы для генеральной совокупности, множество всех элементов которой не известно, на самом деле не делаются, т. к. верификация прогнозов и моделей всегда осуществляется только по выборке из генеральной совокупности. Фактически задача количественного статистического моделирования и прогнозирования в наиболее общем виде формулируется так: необходимо на основании результатов анализа некоторой базовой выборки (или некоторой совокупности таковых) получить модель, в достаточной с точки зрения исследователя степени достоверную и эффективную для описания верифицирующей выборки (или некоторой совокупности таковых) при условии случайности выборок и независимости результатов измерений друг от друга.

Таким образом, для эффективного применения методов прикладной статистики при анализе данных и построении прогнозов можно сформулировать концепцию относительной репрезентативности (оцениваемой количественно на отрезке  $[0;1]$ ) статистической выборки: базовая выборка из генеральной совокупности является репрезентативной по некоторому критерию относительно верифицирующей выборки в том случае, если значение соответствующего критерия, рассчитанное на основании статистического анализа базовой и верифицирующей выборок, принимает экстремальное значение.

При этом качество статистических моделей, формируемых на основании базовой выборки, тем выше, чем мощнее множество верифицирующих выборок, для которых базовая выборка является репрезентативной, при условии минимизации ресурсоемкости получения базовой выборки (минимизации объема базовой выбор-

ки и мощности множества базовых выборок, на основании анализа которых формируется статистическая модель).

Важно отметить, что вышеуказанная концепция в представленном виде относится преимущественно к анализу стационарных случайных величин. При анализе случайных процессов (в особенности, нестационарных) она требует существенных дополнений и уточнений (например, может потребоваться оценка результатов динамического статистического анализа по большому количеству повторных экспериментов), которые рассматриваются авторами как задачи дальнейших исследований.

Существует множество подходов и критериев [8–24], которые применяются при оценке репрезентативности выборки относительно генеральной совокупности (такую теоретическую репрезентативность мы будем называть абсолютной). Все существующие критерии связаны с необходимостью вычисления каких-либо статистических показателей по самой выборке. При этом очевидно, что мощность критериев, опирающихся на этот принцип, существенно меньше, чем мощность критериев, опирающихся на сравнение эмпирических функций плотностей и эмпирических функций распределений вероятности, т. к. при всех дополнительных вычислениях статистических показателей существенно падает число степеней свободы. Главным достоинством подхода, связанного со сравнением эмпирических распределений, является отсутствие необходимости обоснования выбора теоретического распределения, предположительно характеризующего генеральную совокупность.

Результатом сравнения эмпирических распределений является количественная оценка относительной репрезентативности выборки. Применение концепции относительной репрезентативности позволяет объединить достоинства частных и общих критериев репрезентативности, т. к. относительно характера распределения данных генеральной совокупности никаких предположений не делается, значения никаких «предположительных» критериев не вычисляются, а осуществляется построение и критериальное сравнение эмпирических распределения базовой и верифицирующей выборок.

Одним из наиболее «близких к действительности», с нашей точки зрения, критериев репрезентативности является площадь совпадения гистограмм эмпирических распределений базовой и верифицирующей выборок. Однако для обеспечения простоты расчетов и в соответствии с укоренившейся традицией можно для поиска объема репрезентативной выборки применять и сумму квадратов отклонений относительных частот, фактически используя метод наименьших квадратов. Очевидно, что максимум совпадения площадей гистограмм эмпирических распределений соответствует минимуму суммы квадратов их взаимных отклонений.

## 2. Данные для вычислительного эксперимента

Для обобщенного описания структуры экспериментальных данных, подвергаемых анализу и обработке в рамках статистического подхода, в большинстве случаев достаточно модели многомерного тензора, проектируемого на пространство переменных, в которых осуществляется постановка задачи и формулируются результаты статистического анализа. Поскольку визуализация более чем трехмерного пространства вызывает определенные затруднения, мы рассмотрим пример анализа данных, для описания которых достаточно традиционной трехмерной декартовой системы координат.

В качестве анализируемых данных будем рассматривать таблицу значений яркостей пикселей изображения, формируемого сканирующим электронным микроскопом при исследовании субмикрорельефа (топографической морфологии) поверхности модифицированной методом газофазного сульфирования пленки полиэтилена низкой плотности (рис. 1).

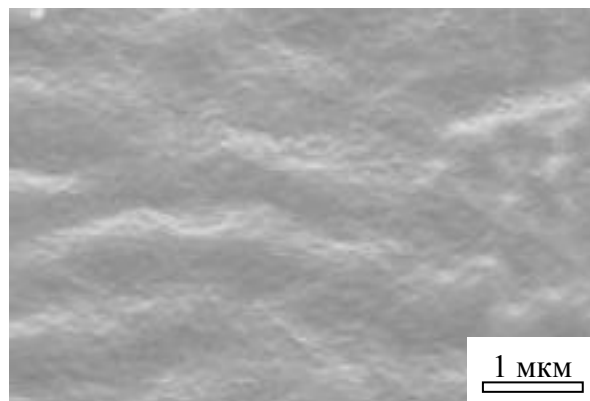


Рисунок 1. Исследуемый фрагмент СЭМ-изображения поверхности образца ПЭНП, модифицированного [25] методом газофазного сульфирования [26]

При визуализации этих данных в декартовой системе координат по осям абсцисс и ординат будут откладываться координаты пикселя, а по оси аппликат — его яркость. Несмотря на то что при описании реального микрорельефа нет никаких оснований полагать, что соответствующие координаты принимают только целочисленные значения, процедура измерений с помощью сканирующего электронного микроскопа автоматически производит дискретизацию измерений, вследствие чего при визуализации по осям системы координат откладываются только натуральные числа (номера столбцов и строк таблицы данных по осям абсцисс и ординат, а также значения уровней яркости пикселей из диапазона  $0 \dots 255$  по оси аппликат). Одним из удобств этого примера является то, что это типичная ситуация, ко-

гда для описания, вообще говоря, непрерывных переменных используется ограниченное количество дискретных значений этих переменных (результатов экспериментальных измерений или наблюдений).

Удобно также и то, что основной задачей статистического анализа таких СЭМ-изображений является выявление на исследуемых поверхностях субмикроразмерных структур (регулярных или стохастических [27, 28]), заведомо меньших по размерам, чем все поле зрения прибора, но обеспечивающих типичные значения тех или иных физико-химических и/или функциональных свойств исследуемых материалов (Что такое «молекула» согласно первоначальному определению? Это такая мельчайшая часть вещества, которая характеризуется той же совокупностью химических свойств, что и вещество в целом. С учетом современных уточнений в нашем случае речь скорее о чем-то вроде домена или макромолекулы, т. к. для отдельных молекул существуют многочисленные размерные эффекты, возникающие при анализе химических реакций на мономолекулярном уровне, а вот «крупные» (десяти тысяч и более) конгломераты молекул уже «ведут себя» как вещество в целом).

В терминах, определения которых были предложены во введении, генеральной совокупностью является вся совокупность возможных СЭМ-изображений поверхностей исследуемого класса материалов (т. е., как и положено, в принципе недостижимо большой объем измерений), верифицирующей выборкой является таблица значений яркостей пикселей, описывающая одно конкретное анализируемое СЭМ-изображение, а возможные базовые выборки формируются как подмножества верифицирующей выборки (т. е. являются фрагментами рассматриваемого СЭМ-изображения). При этом выявление объема базовой выборки, характеризующегося таким же эмпирическим распределением яркостей пикселей, которым характеризуется и верифицирующая выборка, позволяет оценить характерные размеры регулярных или стохастических структур, вносящих определяющий вклад в формирование тех или иных физико-химических и функциональных свойств всего исследуемого материала. Таким образом, абстрактной математической задаче о поиске способа оценки степени репрезентативности выборки сопоставляется конкретная область применения в химии и материаловедении (решается важный этап формирования структурно-функциональной модели материала, связанный с оценкой характерных размеров структурных доменов поверхности материала).

### **3. Результаты вычислительного эксперимента**

Для СЭМ-изображения, представленного на рис. 1, были получены гистограммы эмпирических распределений яркостей пикселей для серии из 40 базовых выборок (на рис. 2 представлены типичные гистограммы в двух равноценных формах их

представления) и гистограмма эмпирического распределения яркостей пикселей для всего анализируемого изображения (рис. 3).

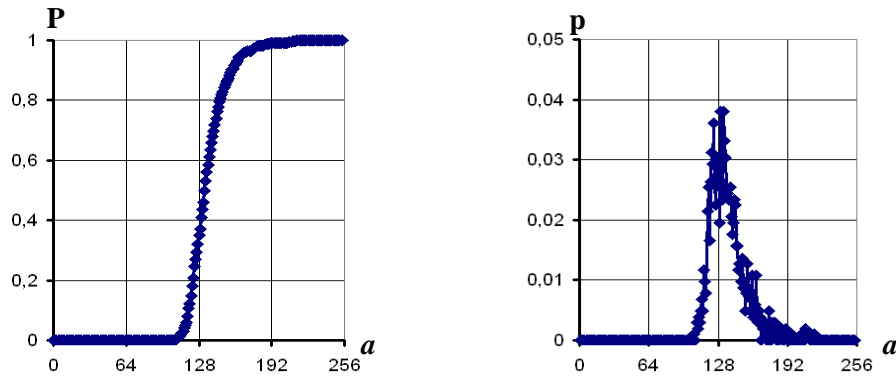


Рисунок 2. Функции распределения и плотности распределения вероятности, характеризующие яркость пикселей базовой выборки объема  $32 \times 32$  элемента

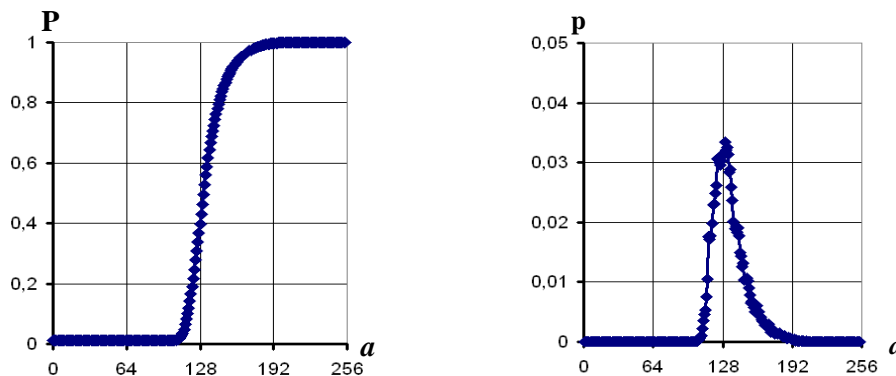


Рисунок 3. Функции распределения и плотности распределения вероятности, характеризующие яркость пикселей верифицирующей выборки объема  $220 \times 152$  элемента

Поскольку при анализе репрезентативности мы хотим использовать традиционные для параметрической математической статистики критерии типа минимизации суммы квадратов отклонений (или суммы модулей отклонений), работа с (интегральными) функциями распределений нам кажется предпочтительной, хотя можно осуществлять соответствующие расчеты, требуя статистической близости не функций, а плотностей распределения яркостей пикселей.

Для анализа репрезентативности рассмотрим, например, зависимость суммы квадратов взаимных отклонений эмпирических распределений, характеризующих всевозможные базовые и верифицирующую выборки. Тривиальным решением яв-

ляется совпадение базовой и верифицирующей выборок, при котором сумма квадратов отклонений гарантированно равна нулю, однако нас это решение интересует мало, т. к. мы хотим найти такие подмножества рассматриваемого множества, чтобы распределения совпадали как можно лучше, а мощность подмножества была минимальной. Поскольку эмпирическое распределение для верифицирующей выборки известно, а эмпирические распределения для базовых выборок нам помогает считать компьютер, то на рис. 3 приводим уже фрагмент (начало обрезано потому, что совсем нет смысла говорить о распределении для выборки, если в ней меньше  $4 \times 4$  элементов, а конец потому, что понятно, что в конце будет 0, но нужно-то найти локальные экстремумы, соответствующие наименьшим возможным объемам выборки) диаграммы зависимости суммы квадратов отклонений эмпирических распределений базовых и верифицирующей выборок от объемов базовых выборок. В нашем примере базовые выборки формируются «в форме» квадратов, т. к. способ обработки исходного материала не предполагает какой-либо анизотропии, что подтверждается предварительными экспериментами по изучению смачиваемости и других физико-химических свойств рассматриваемой поверхности. Объем выборки в этом случае удобно нумеровать количеством пикселей, из которых «состоит» сторона соответствующего квадрата.

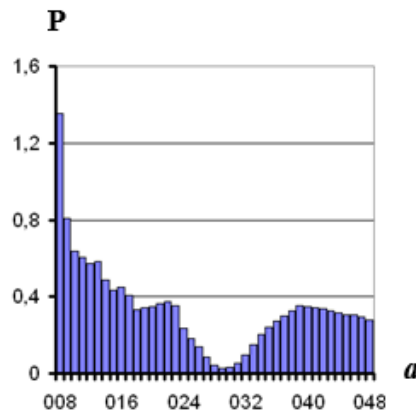


Рисунок 4. Зависимость суммы квадратов отклонений функций распределения вероятности для базовой и верифицирующей выборок от характеристики объема выборки

Прекрасно видно, что выбранный критерий репрезентативности достигает локального минимума при объеме базовой выборки  $29(\pm 1) \times 29(\pm 1)$ , т. е. относительно выборки  $220 \times 152$ , репрезентативной по минимизации отличий функций распределений является уже выборка  $29 \times 29$ , что является оценкой характерных размеров структурного домена рассматриваемой поверхности. Наличие данного локального минимума можно интерпретировать как признак подобия эмпирических

распределений базовой и верифицирующей выборок и считать его выполнением необходимого условия наблюдения фрактальных поверхностных структур.

Косвенным подтверждением необходимости применения специализированных методов выявления фрактальных структур и прямым подтверждение того, что оценка характерного размера структурного домена не является случайной, является наблюдение локального минимума при приблизительно 4-кратном увеличении объема базовой выборки ( $56(\pm 1) \times 56(\pm 1)$ ), репрезентативной по основному критерию метода наименьших квадратов относительно верифицирующей выборки, а также наблюдение соответствующих локальных максимумов при оценке репрезентативности по критерию Колмогорова-Смирнова (по аналогии с [29]). (рис. 5.)

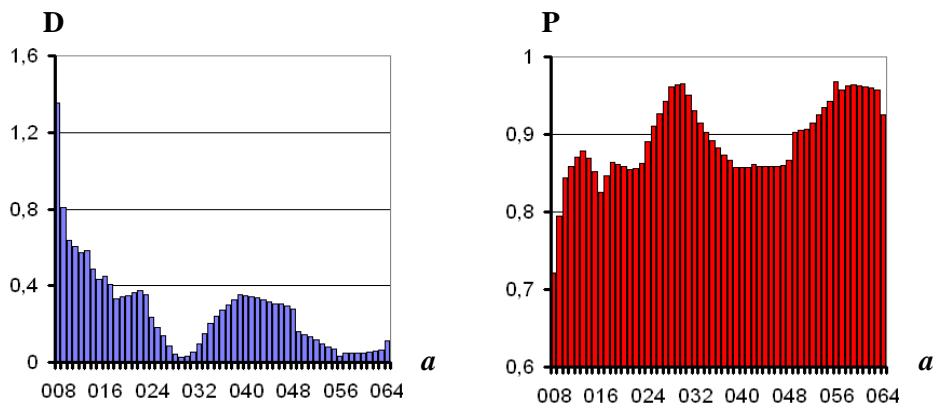


Рисунок 5. Иллюстрация проверки гипотезы о возможности наблюдения фрактальных поверхностных структур. Слева — зависимость суммы квадратов отклонений распределений, характеризующих базовую и верифицирующую выборки; справа — зависимость определяемой на основе критерия Колмогорова-Смирнова достоверности базовой выборки относительно верифицирующей от параметра объема выборки

Анализ результатов вычислительного эксперимента на основе применения критерия Колмогорова-Смирнова показывает, что достоверность базовой выборки относительно верифицирующей изменяется в пределах от 0,73 до 0,96 (при изменении параметра, характеризующего объем выборки, в диапазоне [8; 64]) и достигает локальных максимумов при объемах выборки  $30(\pm 1) \times 30(\pm 1)$ ,  $56(\pm 1) \times 56(\pm 1)$ , что свидетельствует о совпадении интервальных оценок объема репрезентативной выборки двумя разными способами в пределах статистической погрешности.

Неточное совпадение значений удвоенной характеристики объема репрезентативной выборки и характеристики объема выборки, 4-кратно увеличенного относительно репрезентативной объема, обусловлено погрешностями оцифровки изображений и особенностями примененных статистических критериев. Соответствующую



щие функция и плотность распределения яркостей пикселей представлены на рис. 6.

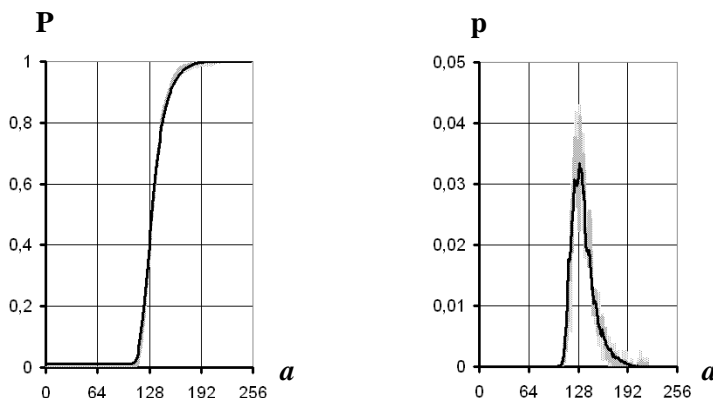


Рисунок 6. Сравнение распределений яркостей пикселей репрезентативных («светло-серая линия» —  $29 \times 29$ , «темно-серая линия» —  $56 \times 56$ ), и верифицирующей («черная линия» —  $220 \times 152$ ) выборок

Для снятия обсуждаемой проблемы, по всей видимости, необходимо дополнительное рассмотрение структурного домена на СЭМ-изображениях с бóльшим увеличением.

Вообще говоря, можно перебрать и другие способы организации базовой выборки (мы рассматривали квадрат «из левого верхнего угла»; можно выбирать прямоугольники, можно — квадраты в разных «локациях» верифицируемой выборки, можно вводить усредненные распределения по ансамблю выборок одинакового объема и т. д.). В общем случае «форма» исследуемой подобласти изображения должна определяться из физических соображений (например, на основе рассмотрения симметрий или на основе наличия априорной информации о каких-либо проявлениях анизотропии) и в многомерных задачах после фиктивного устранения анизотропии с помощью преобразований масштабирования может рассматриваться как гиперкуб, однако этот вопрос не является предметом данной статьи и будет исследован при дальнейшей работе.

Важно то, что появляется инструмент по количественному оцениванию и выбору относительной репрезентативности выборки, а также то, что побочным эффектом является оценка характерных размеров структурных доменов на анализируемых изображениях.

## 4. Заключение

В статье представлена концепция относительной репрезентативности выборки и предложен способ количественной ее оценки. В качестве критерия репрезентативности предлагается использовать площади совпадения гистограмм эмпирических распределений функций вероятностей базовой и верифицирующей выборок, что позволит оценивать степень репрезентативности базовой выборки относительно верифицирующей по предложенному нами численному критерию.

Примером применения и демонстрации физического смысла представленной концепции является анализ полученного средствами сканирующей электронной микроскопии (СЭМ) изображения поверхности газофазно-сульфированной пленки полиэтилена низкой плотности (ПЭНП). Показано, что относительно репрезентативной базовой выборкой для анализируемого изображения является подмножество  $29(\pm 1) \times 29(\pm 1)$  пикселей (с величиной оценки репрезентативности  $0.96 \pm 0.02$ ), что позволяет оценить характерные размеры структурного домена рассматриваемой поверхности.

В рассмотренном примере максимальная площадь совпадения гистограмм эмпирических распределений базовых ( $29 \times 29$  и  $56 \times 56$ ) и верифицирующей ( $220 \times 152$ ) выборок (в обоих случаях) составила  $0.96 \pm 0.02$ , что свидетельствует о высокой степени относительной репрезентативности этих базовых выборок. Минимальная площадь совпадения гистограмм для базовой ( $8 \times 8$ ) и верифицирующей ( $220 \times 152$ ) выборок составила 0.73, что интерпретируется как недостаточная относительная репрезентативность этой ( $8 \times 8$ ) выборки.

Оценка характерных размеров подобластей изображений, характеризующихся эмпирическим распределением яркостей пикселей, близким к распределению яркостей пикселей для всего рассматриваемого изображения, позволяет предположить возможность наблюдения фрактальных структур и указывает на необходимость применения специализированных, более чувствительных и ресурсоемких методов выявления таких структур при дальнейших исследованиях.

## Литература

- [1] Кендалл М. Дж., Стюарт А. Теория распределений / пер. с англ. яз. В. В. Сазонова, А. Н. Ширяева; под ред. А. Н. Колмогорова. — М. : Наука, 1966.
- [2] Кендалл М. Дж., Стюарт А. Статистические выводы и связи / пер. с англ. яз. Л. И. Гальчука, А. Т. Терехина; под ред. А. Н. Колмогорова. — М. : Наука, 1973.

- [3] *Кендалл М. Дж., Стюарт А.* Многомерный статистический анализ и временные ряды / пер. с англ. яз. Э. Л. Пресмана, В. И. Ротаря; под ред. А. Н. Колмогорова. — М. : Наука, 1976.
- [4] *Королюк В. С., Портенко Н. И., Скороход А. В., Турбин А. Ф.* Справочник по теории вероятностей и математической статистике. — М. : Наука, 1985.
- [5] *Вентцель Е. С., Овчаров Л. А.* Теория вероятностей и ее инженерные приложения. — М. : Высшая школа, 2000.
- [6] *Горяинов В. П., Павлов И. В., Цветкова Г. М., Тескин О. И.* Математическая статистика. — М. : Изд-во МГТУ им. Н. Э. Баумана, 2001.
- [7] *Петрушин В. Н., Ульянов М. В.* Информационная чувствительность компьютерных алгоритмов. — М. : ФИЗМАТЛИТ, 2010.
- [8] *Уразбахтин А. И., Уразбахтин И. Г.* Алгоритм проверки однородности выборки и ее репрезентативности исследуемому случайному процессу // *Инфокоммуникационные технологии*. 2006. Т. 4, № 3. С. 10–14.
- [9] *Банах В. А., Смалихо И. Н., Пичугина Е. Л., Брюер А.* Репрезентативность измерений скорости диссипации энергии турбулентности сканирующим когерентным доплеровским лидаром // *Оптика атмосферы и океана*. 2009. Т. 22, № 10. С. 966–972.
- [10] *Мельников Б. Ф., Пивнева С. В., Рогова О. А.* Репрезентативность случайно сгенерированных недетерминированных конечных автоматов с точки зрения соответствующих базисных автоматов // *Стохастическая оптимизация в информатике*. 2010. Т. 6, № 1–1. С. 74–82.
- [11] *Никольчев Е. В., Петрушин В. Н., Малыгин Е. О., Трубочкин А. А., Чертихина И. А.* Вероятностно-интервальный подход к анализу данных // *Вестник Рязанского государственного радиотехнического университета*. 2011, № 38. С. 112–115.
- [12] *Ильясов Ф. Н.* Репрезентативность результатов опроса в маркетинговом исследовании // *Социологические исследования*. 2011, № 3. С. 112–116.
- [13] *Головешкин В. А., Петрушин В. Н., Ульянов М. В.* Количественные оценки информационной чувствительности алгоритмов // *Информационные технологии и вычислительные системы*. 2011, № 4. С. 45–57.
- [14] *Глинский В. В., Гусев Ю. В., Овечкина Н. И., Шмарихина Е. С.* О проверке фактической репрезентативности социальной выборки // *Вестник НГУЭУ*. 2012, № 2. С. 84–90.
- [15] *Таранец И. П., Смуров А. В., Кузнецова Н. А.* Репрезентативность проб разного размера при учете численности и оценке горизонтального распределения почвенных коллембол // *Вестник Московского университета. Серия 16: Биология*. 2012, № 3. С. 44–48.
- [16] *Никольчев Е. В., Петрушин В. Н., Ульянов М. В.* Интервальная оценка средних значений случайной величины в условиях неопределенности функции плотности распределения

- вероятностей // *Известия высших учебных заведений. Проблемы полиграфии и издательского дела*. 2013, № 2. С. 53–59.
- [17] *Дмитриев Я. В.* Определение ошибки репрезентативности и требуемых объемов выборки при расчете значений основных показателей корреляционного, дисперсионного и регрессионного анализа с использованием формулы Чебышева // *Вестник Московского государственного университета приборостроения и информатики. Серия: Приборостроение и информационные технологии*. 2014, № 53. С. 94–100.
- [18] *Петрушин В. Н., Рытиков Г. О., Королев Д. А.* Моделирование жизненного цикла книжной продукции // *Известия высших учебных заведений. Проблемы полиграфии и издательского дела*. 2014, № 2. С. 160–167.
- [19] *Рогова О. А., Софонова Н. В.* Репрезентативность случайно сгенерированных дискретных структур на примерах конечных автоматов и булевых функций // *Эвристические алгоритмы и распределенные вычисления*. 2014. Т. 1, № 6. С. 58–75.
- [20] *Rytkov G. O., Ulianov M. V., Petrushin V. N.* Double smoothing in time series formalization // 2014 International conference on computer technologies in physical and engineering applications (ICCTPEA) Ed. E. I. Veremey. — Санкт-Петербургский государственный университет; IEEE (IEEE Catalog number CFP14BDA-USB). 2014. С. 150–151.
- [21] *Светлова Т. П.* Репрезентативность метеорологических станций и ее влияние на разномасштабные характеристики климатического режима // *Труды Главной геофизической обсерватории им. А. И. Воейкова*. 2015, № 579. С. 115–128.
- [22] *Зубарева В. Е., Петрушин В. Н., Рытиков Г. О.* Особенности применения MS-Excel при прогнозировании социально-экономических временных рядов // *Развитие экономики и менеджмента в современном мире: международная научно-практическая конференция: сборник научных трудов*. 2015. С. 134–137.
- [23] *Иванов Е. Ю., Косяков М. С.* Репрезентативность метрик на основе событий процессора intel sandy bridge при анализе времени обработки данных в памяти // *Программные продукты и системы*. 2015, № 4(112). С. 198–202.
- [24] *Иванов С. П.* Анализ репрезентативности типов леса на арендованной территории // *Актуальные направления научных исследований XXI века: теория и практика*. 2015. Т. 3, № 2-1(13-1). С. 49–52.
- [25] *Назаров В. Г.* Поверхностная модификация полимеров. — М. : Московский государственный университет печати, 2008.
- [26] *Назаров В. Г., Столяров В. П., Молчанов С. П., Юрасик Г. А., Артеменко М. Н.* Поверхностные гетерогенные фторсодержащие макро-, микро- и наноструктуры в полимерных пленках и направления их применения // *Высокомолекулярные соединения. Серия А*. 2013. Т. 55, № 11. С. 1343.

- [27] *Копачев Е. С., Ноздрачев С. А., Петрушин В. Н., Рудяк Ю. В., Рытиков Г. О., Назаров В. Г.* Комплексный метод характеристики изображений поверхностей полимерных композитных материалов // *Физическая мезомеханика*. 2015. Т. 18, № 6. С. 98–110.
- [28] *Petrushin V. N., Rudyak Y. V., Rytikov G. O.* The holistic method of the surface structure characterization // 2016 14th International Baltic Conference on Atomic Layer Deposition, BALD 2016 — Proceedings 14. 2016. С. 15–19.
- [29] *Петрушин В. Н., Дроздов С. А., Рытиков Г. О.* Выявление периодичности и прогнозирование временных рядов в экономике // *Cloud of Science*. 2015. Т. 2, № 2. С. 247–264.

**Авторы:**

*Владимир Николаевич Петрушин* — кандидат физико-математических наук, доцент, доцент кафедры «Менеджмент в сфере культуры и искусства», Институт современного искусства

*Юрий Владимирович Рудяк* — доктор физико-математических наук, профессор, профессор кафедры «Прикладная математика и моделирование систем», Московский политехнический университет

*Георгий Олегович Рытиков* — кандидат технических наук, доцент, старший научный сотрудник Научно-технического центра «Полиграфические и инновационные технологии», Московский политехнический университет

## Representativeness of the statistical sample in the quantitative analysis of images

V. N. Petrushin<sup>\*</sup>, Yu. V. Rudyak<sup>\*\*</sup>, G. O. Rytikov<sup>\*\*</sup>

<sup>\*</sup>Institute of modern art  
Novozavodskaya st., 27A, Moscow, Russia, 121309

<sup>\*\*</sup>Moscow Polytechnic University  
B. Semenovskaya st., 38, Moscow, Russia, 107023

e-mail: petrushinvn@mail.ru

*Abstract.* The article is devoted to the discussion of the concept of "representativeness of a statistical sample". A statistical measure of representativeness and a way of quantitative assessment of the degree of representativeness of a sample from the general population are proposed. To demonstrate the technique of application of the corresponding definitions and methods, an example is considered related to the detection of the volume of a statistically reliable sample in the analysis of SEM images of the surface of a polymer composite.

*Key words:* statistical population, statistical sample, probability density, empirical distribution, histogram, reliability, image analysis.

### References

- [1] Kendall M. Dzh., Styuart A. (1966) *Teoriya raspredeleniy*. Moscow, Nauka. [In Rus]
- [2] Kendall M. Dzh., Styuart A. (1973) *Statisticheskiye vyvody i svyazi*. Moscow, Nauka. [In Rus]
- [3] Kendall M. Dzh., Styuart A. (1976) *Mnogomernyy statisticheskiy analiz i vremennyye ryady*. Moscow, Nauka. [In Rus]
- [4] Korolyuk V. S., Portenko N. I., Skorokhod A. V., Turbin A. F. (1985) *Spravochnik po teorii veroyatnostey i matematicheskoy statistike*. Moscow, Nauka. [In Rus]
- [5] Venttsel' Ye. S., Ovcharov L. A. (2000) *Teoriya veroyatnostey i yeye inzhenernyye prilozheniya*. Moscow, Vysshaya shkola. [In Rus]
- [6] Goryainov V. P., Pavlov I. V., Tsvetkova G. M., Teskin O. I. (2001) *Matematicheskaya statistika*. Moscow, Izd-vo MGTU im. N. E. Bauman. [In Rus]
- [7] Petrushin V. N., Ul'yanov M. V. (2010) *Informatsionnaya chuvstvitel'nost' komp'yuternykh algoritmov*. Moscow, FIZMATLIT. [In Rus]
- [8] Urazbakhtin A. I., Urazbakhtin I. G. (2006) *Infokommunikatsionnyye tekhnologii*, **4**(3): 10–14.
- [9] Banakh V. A., Smalikho I. N., Pichugina E. L., Bryuyer A. (2009) *Optika atmosfery i okeana*, **22**(10):966–972. [In Rus]

- [10] Mel'nikov B. F., Pivneva S. V., Rogova O. A. (2010) *Stokhasticheskaya optimizatsiya v informatike*, **6**(1–1):74–82. [In Rus]
- [11] Nikulchev E. V., Petrushin V. N., Malygin E. O., Trubochkin A. A., Chertikhina I. A. (2011) *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, **38**:112–115. [In Rus]
- [12] Il'yasov F. N. (2011) *Sotsiologicheskiye issledovaniya*, **3**:112–116. [In Rus]
- [13] Goloveshkin V. A., Petrushin V. N., Ul'yanov M. V. (2011) *Informatsionnyye tekhnologii i vychislitel'nyye sistemy*, **4**:45–57. [In Rus]
- [14] Glinskiy V. V., Gusev Yu. V., Ovechkina N. I., Shmarikhina E. S. (2012) *Vestnik NGUEU*, **2**:84–90. [In Rus]
- [15] Taranets I. P., Smurov A. V., Kuznetsova N. A. (2012) *Vestnik Moskovskogo universiteta. Seriya 16: Biologiya*, **3**:44–48. [In Rus]
- [16] Nikul'chev E. V., Petrushin V. N., Ul'yanov M. V. (2013) *Izvestiya vysshikh uchebnykh zavedeniy. Problemy poligrafii i izdatel'skogo dela*, **2**:053–059. [In Rus]
- [17] Dmitriyev Ya. V. (2014) *Vestnik Moskovskogo gosudarstvennogo universiteta priborostroyeniya i informatiki. Seriya: Priborostroyeniye i informatsionnyye tekhnologii*, **53**:94–100. [In Rus]
- [18] Petrushin V. N., Rytikov G. O., Korolev D. A. (2014) *Izvestiya vysshikh uchebnykh zavedeniy. Problemy poligrafii i izdatel'skogo dela*, **2**:160–167. [In Rus]
- [19] Rogova O. A., Sofonova N. V. (2014) *Evristicheskkiye algoritmy i raspredelennyye vychisleniya*, **1**(6):58–75. [In Rus]
- [20] Rytikov G. O., Ulianov M. V., Petrushin V. N. (2014) Double smoothing in time series formalization. In Proc. 2014 International conference on computer technologies in physical and engineering applications (ICCTPEA) Ed. E. I. Veremey. Sankt-Peterburgskiy gosudarstvennyy universitet; IEEE (IEEE Catalog number CFP14BDA-USB). P. 150–151. [In Rus]
- [21] Svetlova T. P. (2015) *Trudy Glavnoy geofizicheskoy observatorii im. A. I. Voyeykova*. **579**:115–128. [In Rus]
- [22] Zubareva V. Ye., Petrushin V. N., Rytikov G. O. (2015) Osobennosti primeneniya MS-Excel pri prognozirovanii sotsial'no-ekonomicheskikh vremennykh ryadov. In proc. Razvitiye ekonomiki i menedzhmenta v sovremennom mire: mezhdunarodnaya nauchno-prakticheskaya konferentsiya. P. 134–137. [In Rus]
- [23] Ivanov E. Yu., Kosyakov M. S. (2015) *Programmnyye produkty i sistemy*, **4**(112):198–202. [In Rus]
- [24] Ivanov S. P. (2015) *Aktual'nyye napravleniya nauchnykh issledovaniy XXI veka: teoriya i praktika*, **3**(2-1):49–52. [In Rus]
- [25] Nazarov V. G. (2008) *Poverkhnostnaya modifikatsiya polimerov*. Moscow, Moskovskiy gosudarstvennyy universitet pechati. [In Rus]

- [26] *Nazarov V. G., Stolyarov V. P., Molchanov S. P., Yurasik G. A., Artemenko M. N.* (2013) *Vysokomolekulyarnyye soyedineniya. Seriya A*, **55**(11):1343. [In Rus]
- [27] *Kopachev E. S., Nozdrachev S. A., Petrushin V. N., Rudyak Yu. V., Rytikov G. O., Nazarov V. G.* (2015) *Fizicheskaya mezomekhanika*, **18**(6):98–110. [In Rus]
- [28] *Petrushin V. N., Rudyak Y. V., Rytikov G. O.* (2016) The holistic method of the surface structure characterization. In Proc. 2016 14th International Baltic Conference on Atomic Layer Deposition, BALD 2016 — Proceedings 14. P. 15–19. [In Rus]
- [29] *Petrushin V. N., Drozdov S. A., Rytikov G. O.* (2015) *Cloud of Science*, **2**(2):247–264. [In Rus]