УДК 519.281

РАЗДЕЛЕНИЕ СМЕСЕЙ РАСПРЕДЕЛЕНИЙ ВЕРОЯТНОСТЕЙ НА ИХ СОСТАВЛЯЮЩИЕ

О. К. Исаенко, В. Ю. Урбах

Работы, рассматриваемые в этом обзоре, посвящены следующей задаче: известно из априорных соображений, что заданная совокупность объектов является смесью некоторого (известного или неизвестного) числа отдельных совокупностей, и требуется найти эти совокупности. Под нахождением составляющих совокупностей следует понимать (1) оценивание их параметров или (2) построение решающего правила для отнесения элементов к подсовокупностям. Решив задачу (1), можно затем по известным методам (дискриминантный, байесовский и др.) построить решающее правило, т. е. решить задачу (2). С другой стороны, решив сначала задачу (2), можно после разнесения элементов по подсовокупностям оценить параметры распределений в этих подсовокупностях. Решение задачи разделения смеси в порядке (1)—(2) требует использования определенной информации о типах распределений в совокупностях, входящих в смесь. В этом случае могут применяться известные классические методы, такие как метод моментов, метод максимального правдоподобия, метод наименьших квадратов и др. Если указанная информация отсутствует или почему-либо нежелательно ее использовать, например, из-за вычислительных трудностей, то приходится идти путём (2)—(1), приводящим к применению методов так называемого кластер-анализа (cluster—группа, рой).

Идентифицируемость смеси. Пусть имеется параметрическое семейство функций распределения $\mathcal{F}=\{F(x, \alpha);\ \alpha \in R_1^m$, где R_1^m — борелевское подмножество $\subset R^m$, $x \in R^n\}$. Обозначим через $\mathcal{G}$ — семейство функций распределения $G(\alpha)$. Определим $Q(G)$ как

$$Q(G)=\int\limits_{R_1^m} F(x, \alpha)\, dG(\alpha)$$

или

$$Q(G)=\sum_{i=1}^{N} F(x,\ \alpha_i)\cdot\nu_i, \text{ где } \sum_{i=1}^{N}\nu_i=1.$$

В первом случае имеем непрерывную G-смесь функций распределения $F(x,\ \alpha)$; $G(\alpha)$ называется смешивающим распределением, которое и требуется найти. Во втором случае имеем конечную смесь функций распределения $F(x,\ \alpha_i)$, и ставится задача нахождения весовых коэффициентов ν_i и параметров α_i составляющих совокупностей (функции $F(x,\ \alpha_i)$ могут и не принадлежать одному и тому же типу).

Скажем, что множество непрерывных смесей $\mathcal{F}$ идентифицируемо, если Q имеет обратную функцию, и для конечной смеси $\mathcal{F}$ идентифицируемо тогда и только тогда, когда из

$$\sum_{i=1}^{N}\nu_i F(x,\ \alpha_i)=\sum_{j=1}^{M}\nu'_j F(x,\ \alpha'_j)$$

следует, что $N=M$, и для каждого $i(1\leqslant i\leqslant N)$ имеется некоторое $j(1\leqslant j\leqslant M)$ такое, что $\nu_i=\nu'_j$ и $\alpha_i=\alpha'_j$. Очевидно, что однозначное решение задачи разделения смеси существует, только если $\mathcal{F}$ есть идентифицируемая смесь. Систематическое изучение проблемы идентифицируемости смесей проведено в [79, 80, 81]. В [81] автор дает достаточные условия идентифицируемости конечных смесей и доказывает, что одномерные гауссовские, гамма-распределения и пуассоновские распределения идентифицируемы. Расширение этой работы содержится в [90], где показана идентифицируемость семейства p-мерных нормальных распределений, семейства, составленного из p произведений одномерных экспоненциальных распределений и композиции последних двух семейств; для одномерных распределений это верно также для семейства распределений Коши, невырожденных членов отрицательно биномиального семейства и семейства сдвигов любой одномерной функции распределения и, кроме того, для семейства n произведений одномерных идентифицируемых распределений [82]. Вопрос идентифицируемости смесей рассмотрен также в [21, 30, 59]. В дальнейшем идентифицируемость смесей будет подразумеваться.

Метод моментов. Большинство работ посвящено решению одномерной задачи разделения смесей. Впервые она была сформулирована Пирсоном (1897) [63], применившим метод моментов (м. м.). Этот метод заключается в том, что к моментам смеси (выборочным или заданным) приравниваются теоретические моменты смеси, выраженные через теоретические моменты составляющих совокупностей и веса составляющих совокупностей. Для простоты можно ограничиться числом уравнений, равным числу оцениваемых параметров, и рассматривать уравнения только для низших моментов. Для случая смеси двух распределений

$$f(x)=\nu\cdot f_1(x)+(1-\nu)\cdot f_2(x),\quad 0<\nu<1, \tag{1}$$

r-ый начальный момент смеси получается равным

$$m_r=\nu\cdot\int_{-\infty}^{+\infty}x^r f_1(x)\,dx+(1-\nu)\cdot\int_{-\infty}^{+\infty}x^r\cdot f_2(x)\,dx. \tag{2}$$

Если предположить, что $f_1(x)=N(\mu_1, \sigma_1)$, $f_2(x)=N(\mu_2, \sigma_2)$, то для нахождения пяти неизвестных параметров $(\mu_1, \mu_2, \sigma_1, \sigma_2, \nu)$ можно ограничиться первыми пятью уравнениями, т. е. положить $r=0, 1, \ldots, 4$, и поставленная задача становится чисто алгебраической. В частности, в работе Пирсона она сводилась к решению уравнения девятой степени. В [29] предложен метод, облегчающий решение этого уравнения. Автор предлагает находить параметр $u=\mu_1+\mu_2$ при максимизации функции правдоподобия

$$L(u)=\prod_{i=1}^{N}f(x_i, \bar{x}, m_2, m_3, m_4)$$

(остальные искомые параметры выражаются через u и $v=\mu_1\cdot\mu_2$) или при минимизации

$$\chi^2(u, \bar{x}, m_2, m_3, m_4)=\sum_{i=1}^{k}\frac{(n_i-l_i)^2}{l_i},$$

где $l_i=N\cdot p_i$ — ожидаемое число наблюдений в i-ом интервале, первые четыре выборочных момента приравнены к соответствующим теоретическим моментам. Такая процедура предложена автором вследствие того, что оценки по м. м. п. для пяти неизвестных параметров находить трудно, а при м. м. они прямо выражаются через u и v. Автор приводит иллюстративный пример с $N=1000$ и предлагает использовать выведенные оценки при больших выборках, вычисляя выборочные моменты из негруппированных данных. Бюрро и Стрёмгрен [24, 75] решали такую же задачу разделения смеси двух нормальных распределений с помощью метода номограмм. В ряде работ [8, 9, 13, 49] рассмотрены частные случаи этой задачи, когда имеется та или иная информация о параметрах составляющих распределений (например, $\nu=\frac{1}{2}$, или $\sigma_1=\sigma_2$ и т. п.). Графическое решение этих уравнений дано в [35] для двух распределений Вейбулла. Автор пытался применить к этой смеси м. м. п., метод факториальных моментов, но нашел, что описанный им метод лучше, так как дает хорошее согласие с истинными оценками (проверено на примерах). В [67, 68] используется метод моментов для нахождения оценок параметров в смеси двух экспоненциальных распределений, пуассоновского, биноминального и распределений Вейбулла. В [28] метод, примененный в последних работах для оценки параметров смеси двух пуассоновских распределений, упрощается за счет использования факториальных моментов, а также там получены оценки параметров в смеси усеченных пуассоновских распределений и в случае смеси пуассоновского и биномиального распределения. В [50] метод моментов применяется для разделения смеси двух гамма-популяций, а в [54] для разделения любого конечного числа экспоненциальных распределений. Смесь некоторого числа биномиальных распределений в деталях исследована в [19, 20], где сначала показана идентифицируемость смеси при некоторых ограничениях на параметры. Имеются работы, в которых исследуются некоторые свойства выборочных оценок параметров. Так, в [20] конструируются наилучшие асимптотически нормальные оценки, основанные на моментных оценках. Показывается, что оценки параметров по методу моментов состоятельны, но не эффективны, хотя в некоторых частных случаях имеется асимптотическая относительная эффективность оценок. О низкой эффективности моментных оценок для простейших случаев говорится в [77]. Состоятельность моментных оценок освещена в [6, 31] и в большинстве других работ, использующих этот метод. В [49, 50, 72] доказана асимптотическая нормальность оценок. Асимптотическая дисперсия для оценок параметров в смеси распределений Максвелла найдена в [85], в [67] находятся дисперсии оценок параметров в смеси экспоненциальных распределений при известных весах. В [23] описано численно пространство приемлемых решений для смеси двух нормальных распределений в терминах трех стандартных семиинвариантов K_3, K_4, K_5 при моментном подходе к оцениванию пяти неизвестных параметров. Общее заключение о применимости моментных оценок, которое показывает, что в одномерном случае и при больших Δ (расстояние между популяциями) моментные оценки, хотя и не являются эффективными, могут использоваться при практических расчетах, дано Деем [31]. Необходимо отметить, что уравнения для моментов некоторыми авторами выписываются из других соображений, чем дано выше. Например, возможны такие модификации: в [47, 76] решаются уравнения для дробных моментов, в [13, 53] (в последней работе рассматривается смесь распределений Лапласа) — для абсолютных моментов, в [19, 20, 28] — для факториальных моментов.

Необходимо отметить неоднозначность решения системы уравнений при нахождении оценок параметров по методу моментов [43, 52]. В последних работах предложены добавочные критерии для выбора одного из решений.

Некоторое эмпирическое сравнение метода моментов и графического проведено в [60] для смеси двух нормальных распределений. Применение метода моментов к многомерным задачам рассматривалось в [3, 49] только для случая нормальных распределений, причем результаты ограничиваются лишь записью системы уравнений.

Метод максимума правдоподобия. Более полно и широко рассмотрено в литературе решение обсуждаемой задачи методом максимального правдоподобия (м. м. п.). Во всех работах, относящихся к этому методу, развиваются те или иные итерационные схемы, приводящие к максимуму функцию правдоподобия. Уравнения правдоподобия преобразуются к виду

$$\theta_j = \Phi(\theta_{j-1}), \tag{3}$$

где θ — искомый набор параметров, а j — номер итерации. В [12] предлагается алгоритм самообучения, когда распределения вероятностей каждого образа принадлежат к параметрическому семейству. Алгоритм вычисляет по некоторым значениям параметров новые, более правдоподобные значения. Доказаны теоремы о том, что независимо от начальных значений параметров алгоритм обеспечивает сходимость к оценкам максимального правдоподобия. Сообщается, что проводилась экспериментальная проверка для случая большого числа неизвестных параметров и что алгоритм работает хорошо, но конкретные результаты не приводятся.

Для смеси M нормальных популяций этот метод дан в [16, 31, 41, 49]. В [41] строятся две итерационные схемы: метод скорейшего спуска и метод Ньютона и обсуждаются возможности для их сходимости. При реализации этих схем видно, что метод скорейшего спуска всегда увеличивал значения функции правдоподобия и хорошо работал при малом числе данных. Ньютоновская итерационная схема хорошо применима при большом объеме выборки, кроме того, оценки, полученные по этой схеме, оказываются более близкими к истинным параметрам модели. Во всех примерах использовался объем выборки, больший 200. Чтобы охарактеризовать трудности проблемы, автором были подсчитаны теоретические асимптотические дисперсии оценок весов (при других известных параметрах). Асимптотические дисперсии оценок для весов возрастают очень сильно с уменьшением расстояния между средними.

Нахождение параметров в смеси M экспоненциальных распределений м. м. п. производится в [42]. Сходимость предложенной итерационной процедуры не доказывается, но показано, что после первых ста итераций дальнейшее увеличение их числа не меняет заметным образом значение функции правдоподобия. На IBM 360—75 для исследуемых 207 выборок из примеров потребовалось 5 мин. Желательный объем выборки не менее 1000. М. м. п. для смеси двух гамма-популяций схематично рассмотрен в [50].

Возможности применения м. м. п. для оценки параметров смеси при малом объеме выборки исследованы в [45]. Асимптотические дисперсии для оценок параметров смеси подсчитаны в [31, 41, 42, 71]. При сравнении м. м. и м. м. п. во всех работах предпочтение отдается м. м. п. как наиболее эффективному. Например, это отмечено в [37, 78]. В последней из упомянутых работ эффективность м. м. относительно м. м. п. определяется как отношение детерминантов асимптотических ковариационных матриц. Моментные оценки могут использоваться в качестве начальных точек в итерационной процедуре, хотя, как выводится в [31], при малых Δ моментные оценки могут оказываться настолько плохими, что непригодны даже в качестве начальных точек.

Преимущества м. м. п., по сравнению с м. м., проявляются главным образом в многомерных задачах. Однако можно указать только на три работы, в которых рассматриваются подобные задачи, причем во всех случаях распределения в составляющих совокупностях считаются нормальными.

В [31] уравнения (3) решаются известными итерационными методами. Строгого доказательства сходимости итерационной процедуры не дано; утверждается существование решения (возможно, не единственного), однако доказательства этого утверждения не приводится. При малых значениях $\Delta (\Delta < 2)$ и выборочном объеме меньшем, чем 100, и при размерности, большей 3, существует несколько решений за счет появления случайных сгущений. Тогда предлагается оценивать функцию правдоподобия у каждого локального максимума, чтобы найти главный максимум. При повторении итерационной процедуры из достаточного числа различных начальных точек мы можем найти все локальные максимумы. Скорость сходимости итерационной процедуры зависит от расстояния между совокупностями и выбора начального приближения θ_0. В случае смеси двух нормальных совокупностей с равными ковариационными матрицами м. м. п. позволяет получить достаточно «хорошие» оценки параметров составляющих совокупностей, особенно, когда размерность выборочного пространства меньше 10. Рассмотрено расширение метода для смеси более двух совокупностей. Некоторое обобщение этих результатов дано в [87].

Применение м. м. п. в [49] имеет несколько иную форму, чем в остальных работах. Функция правдоподобия выписывается в виде

$$\lambda = (2\pi)^{-N \cdot p/2} |\Sigma|^{-N/2} \cdot \exp\left[-\sum_{i=1}^{2} \sum_{r=1}^{N} \theta_{ir} (x_r - \mu_i)^T \Sigma^{-1} (x_r - \mu_i)/2 \right] \quad (4)$$

и максимизируется надлежащим выбором θ_{ir}. Параметры μ_1, μ_2, Σ выражаются через $\hat{\theta}_{ir}$ (значение θ_{ir}, при котором функция правдоподобия максимальна) равенствами:

$$\mu_i=\sum_r \hat{\theta}_{ir}x_r \Big/ \sum \hat{\theta}_{ir}, \quad i=1,\ 2,$$

$$\Sigma=\sum_i \sum_{r=1}^{N} \hat{\theta}_{ir}(x_r-\mu_i)(x_r-\mu_i)^T/N. \tag{5}$$

Для нахождения $\hat{\theta}_{ir}$ предложена итерационная процедура, сходимость которой не доказана. Рассмотрен иллюстративный пример для смеси двумерных нормальных плотностей, в котором получены оценки параметров, близкие к истинным значениям. Утверждается (без доказательства), что выборочные свойства оценки для θ_{ir}, полученной таким образом, являются удовлетворительными, если Δ не слишком мало. Для того чтобы получить выборочные моменты оценок по м. м. п. для других параметров, необходимо проводить дальнейшие вычисления.

В третьей работе [70] рассматривается следующая модель. Кроме $\mu_1, \ldots, \mu_M$, $\Sigma_1, \ldots, \Sigma_M$, ищется параметр группирования $\gamma=(\gamma_1, \ldots, \gamma_N)$, где γ_g равно g, если элемент выборки взят из g-й совокупности. Если $\theta=(\gamma, \mu_1, \ldots, \mu_M, \Sigma_1, \ldots, \Sigma_M)$, то логарифмическая функция правдоподобия даётся формулой

$$l(\theta)=-\frac{1}{2}\cdot\sum_{g=1}^{M}\Big(\sum_{G_g}(y_i-\mu_g)^T\Sigma^{-1}(y_i-\mu_g)+n_g\cdot\log|\Sigma_g|\Big),$$

где G_g — множество элементов выборки, отнесённых в g-ую группу, n_g — число элементов в G_g, $|\Sigma_g|=\det\Sigma_g$. Основная проблема — оценить N-мерный вектор γ и, следовательно, группы $G_1, G_2, \ldots, G_M$. Оценка $\hat{\gamma}$ по м. м. п. для γ находится при минимизации

$$\sum_{g=1}^{M}\{\operatorname{tr} W_{gy}\Sigma^{-1}+n_g\cdot\log|\Sigma_g|\},$$

где $W_{gy}=\sum_{G_g}(y_i-\bar{y}_g)(y_i-\bar{y}_g)^T$. Далее, если $\Sigma_g=\Sigma\ (g=1, \ldots, M)$ и, считая Σ известной (или имеем достаточно большую выборку для получения «хорошей» оценки для Σ), $\hat{\gamma}$ находится при минимизации

$$\operatorname{tr}(W_y\cdot\Sigma^{-1})=\operatorname{tr}\left(\sum_g W_{gy}\cdot\Sigma^{-1}\right)=\sum_{g=1}^{M}\sum_{G_g}(y_i-\bar{y}_g)^T\Sigma^{-1}(y_i-\bar{y}_g),$$

т. е. минимизируется общая внутригрупповая сумма квадратов. Это есть один из известных экстремальных подходов к разделению совокупности на составляющие подсовокупности. Если Σ не является известной, то $\hat{\gamma}$ есть тот параметр группирования, который минимизирует W_y; этот подход также широко применяется в кластер-анализе. Далее снимается ограничение на равенство ковариационных матриц составляющих подсовокупностей, и тогда $\hat{\gamma}$ ищется при минимизации $\prod_{g=1}^{M} |W_{gy}|^{n_g}$ (необходимо иметь по крайней мере $p+1$ наблюдение из каждой группы, где p — размерность пространства). Приведенные числовые результаты показывают возможность применения указанных подходов. Вопросы качества получаемых оценок в работе не обсуждаются. В [31] и [70] методы работают для случая смеси M многомерных совокупностей, в [49] — только для специального случая $M=2$. Во всех этих работах указывается на непропорциональный рост вычислительных трудностей при увеличении размерности p. Кроме того, необходимо дальнейшее исследование свойств полученных оценок, а также расширение методики на случай, когда не имеется сведений о количестве распределений, входящих в смесь.

Необходимо отметить две работы [7, 37], сравнивающие м. м., м. м. п. и метод минимума χ^2 в применении к задаче разделения смеси. В [7] говорится о трудностях реализации м. м. п., так как система уравнений для м. м. п. имеет решения, соответствующие седловым точкам поверхности L (функция правдоподобия), в пространстве параметров. Кроме того, с ростом N среднеквадратичная погрешность оценок уменьшается со скоростью $\sqrt{N}$, а объем вычислений растет пропорционально N — в то время как м. минимума χ^2, близкий по универсальности и эффективности к м. м. п., не требует увеличения объема вычислений. Для определения θ^* по м. м. χ^2 пространство реализаций должно быть предварительно разбито на $s \ll N$ непересекающихся областей $V_1, V_2, \ldots, V_s$. Искомыми оценками θ^* являются такие значения параметров θ, которые обеспечивают минимум выражения

$$\chi^2 = N \sum_{i=1}^{s} \frac{\left(\nu_i^* - \nu_i(\theta)\right)^2}{\nu_i(\theta)},$$

где ν_i^* — относительное число реализаций, попавших в i-ую область V_i; $\nu_i(\theta)$ — вероятность попадания в эту область, равная

$$\nu_i(\theta) = \int \ldots \int_{V_i} f(x, \theta)\, dx_1 \ldots dx_p.$$

Обсуждается вопрос выбора областей V_i. Описан практический путь получения оценок θ^* для смеси многомерных нормальных распределений, вначале ищущий параметры раздельно для каждой из p компонент. Никаких конкретных данных об объеме вычислений, времени для нахождения параметров не приведено. Для иллюстрации точности получаемых оценок на графиках представлены асимптотическая дисперсия σ_{11} и смешанный второй момент σ_{12} оценок μ_1^* и μ_2^* двух одномерных нормальных распределений с одинаковой априорно известной дисперсией и $\nu_1=\nu_2=0{,}5$. Видно, что м. минимума χ^2 при увеличении s приближается по эффективности к м. м. п. и оба они при увеличении расстояния γ между средними стремятся к пределу, соответствующему режиму обучения. Дисперсия оценок по м. м., начиная с некоторого достаточно малого γ, неограниченно возрастает. В [37] методы моментов, м. п. и м. χ^2 численно сравнены по смещению порядка N^{-1} и по среднеквадратичной ошибке порядка N^{-2} для смеси двух нормальных плотностей. Смесь двух нормальных плотностей имеет две или четыре точки перегиба. Распределения с двумя перегибами имеют только одну моду. Для четырех точек перегиба распределение может быть бимодальным или унимодальным. Авторы сравнили характеристики оценок по этим методам для большого числа смесей всех трех типов, что позволяет им рассматривать результаты как типичные. Показано, что сравниваемые три метода существенно не различаются относительно смещения, но относительно среднеквадратичной ошибки группированные оценки оказываются более аккуратными, чем моментная оценка.

В [46] оценки по м. м. п. сравнены для трех различных типов выборок, моделируемых по методу Монте-Карло. Если выборка из смеси распределений может быть дополнена выборкой небольшого объема (10% от объема всей выборки), элементы которой имеют функцией распределения f_1 или f_2 (составляющие смеси), или если часть выборки из смеси может быть корректно классифицирована, то оценки, получаемые по этим выборкам, более аккуратны, чем оценки параметров, построенные просто по выборке из смеси без какой-либо информации об ее элементах. Оценки по м. м. п. и м. м. для смеси двух нормальных распределений при имеющейся дополнительной информации об одном из распределений, составляющих смесь, рассмотрены в [33].

Метод наименьших квадратов. Если параметры составляющих совокупностей известны и нужно найти только веса этих составляющих совокупностей, то можно применить метод наименьших квадратов [14, 26, 27, 62]. В этом методе веса составляющих совокупностей находятся из условия минимума интеграла квадратичного отклонения заданного распределения смеси от распределения, отвечающего модели смеси с варьируемыми значениями весов. В [62] описан практический путь построения функций, линейная комбинация которых дает искомую оценку для смешивающего распределения. В этой работе, кроме оценки по м. н. кв., авторы предлагают байесовскую оценку для дискретного множества параметров, минимизирующую квадратичную функцию потерь. Показано, что при некоторых определенных условиях эта байесовская оценка сходится в среднеквадратичном и с вероятностью 1 к некоторой асимптотической оценке, скорость сходимости вычислена. Кроме того, определены условия, при которых этот асимптотический вектор оценки является ближайшим к вектору истинных параметров в метрике, определяемой фишеровской информационной матрицей. Рассмотренные алгоритмы численно реализуемы на ЭВМ, хотя автор отмечает (впрочем, без конкретных количественных оценок) увеличение необходимой памяти ЭВМ при росте объема выборки и размерности изучаемых объектов.

В [26] дана оценка весов составляющих совокупностей, применимая как к непрерывным смесям, так и к конечным. Полученная оценка состоятельна и имеет асимптотически нормальное (многомерное) распределение. Результаты, полученные при изучении выборок из смеси M нормальных распределений, моделируемых методом Монте-Карло, показывают, что описанный метод хорошо работает и для выборок малого объема. В приведенной таблице сведены выборочные размеры, для которых распределения оценки оказываются приближенно нормальными. В [27] этот же метод развит для оценки как веса, так и параметров составляющих совокупностей для конечной смеси. Полученные оценки являются состоятельными и асимптотически нормальными. Скорость сходимости этих оценок к истинным значениям параметров есть $O(N^{-1/2})$. Числовые результаты не приводятся.

Другие методы, использующие элементы выборки для оценивания параметров смеси. Задачи, в которых неизвестны только веса составляющих совокупностей, решаются также и другими методами. Минимаксная (минимизируется максимум риска) оценка дана в [22]. В качестве функции риска берется функция, которая является взвешенной суммой диагональных элементов обратной информационной матрицы. Свойства этой оценки подлежат дальнейшему исследованию. В [48, 58] сконструирована состоятельная байесовская оценка для смешивающих распределений в дискретном случае, причем в [48] дано также расширение для непрерывной функции распределения вероятностей смеси (методы получения априорных вероятностных распределений различны). В [32] предложена оценка смешивающего распределения, являющаяся ступенчатой функцией с конечным числом ступенек; в [64] сконструирована оценка смешивающего распределения кусочно-полиномиальными дугами. Исследуется качество полученной оценки для моделированных данных по трем критериям: визуальное сравнение; эксцесс ожидаемого байесовского риска; среднеквадратичная ошибка оценки процентных точек смешивающего распределения, — заключается, что полученная оценка является «хорошей» в смысле введенных критериев. Общий алгоритм для конструирования состоятельной оценки смешивающего распределения дан в [88]. Там же доказывается эффективность этого алгоритма для всех идентифицируемых семейств распределений. В [89] этот же автор излагает свои теоретические результаты в связи с проблемой распознавания образов, но никаких конкретных применений описанного метода не приводится.

В работе [17] доказана некоторая теорема о симметрических статистиках для выборки из смеси двух распределений, позволяющая строить критерии относительно параметра веса смеси. Например, при гипотезе $\nu=0$ выборочное среднее распределено нормально; при гипотезе $0<\nu<1$ $\bar{X}$ является смесью $(n+1)$-го нормального распределения с общей дисперсией $\frac{\sigma^2}{n}$ и со средним $\mu_k=\frac{k}{n}\mu_1+\left(1-\frac{k}{n}\right)\mu_2$. В [25] развит следующий метод оценки весового коэффициента смеси для многомерных распределений. Оценка $\tilde{\nu}$ для ν предлагается в виде отношения линейных комбинаций меток рангов наблюдений. Оптимизация коэффициентов линейных комбинаций меток для минимизации асимптотической дисперсии оценки $\tilde{\nu}$ приводит к оценке $\hat{\nu}$. Доказана состоятельность и асимптотическая нормальность оценки $\hat{\nu}$.

Для разделения смеси на составляющие применяются также графические методы [18, 40]. Интересна работа [91], в которой развивается алгоритм стохастической аппроксимации для смеси одномерных нормальных функций распределения. Алгоритм минимизирует информационный критерий, обсуждаются два варианта этой задачи: а) истинная функция плотности есть линейная комбинация нормальных функций плотности и б) функциональная форма плотности неизвестна, а нормальная смесь используется как аппроксимация. Даны условия сходимости алгоритма, хотя замечено, что он (алгоритм) может сходиться к локальному максимуму. Авторы принимают, что существует априорная информация, позволяющая выбрать начальную оценку, которая разумно близка к глобальному максимуму. Приведен числовой пример, для которого потребовалось около 10 минут на ЭВМ и 5000 итераций. В [51] строится теория оценки параметров p_i и μ_i, $i=1,\ldots,r$, для плотности вероятности вида

$$f(x)=X_\varkappa(x)\sum_i p_i \exp\{B(\mu_i)x+C(x)+D(\mu_i)\},$$

где $X_\varkappa(x)$ — характеристическая функция заданного интервала, B, C, D — известные функции. Показывается состоятельность и асимптотическая нормальность полученных оценок, общая теория применяется к оценке параметров смесей двух биномиальных и смеси двух экспоненциальных распределений. В [83] решается задача разделения смеси в случае однопараметрических распределений, составляющих смесь

$$f(x, \alpha_1,\ldots,\alpha_k)=\sum_i \nu_i f_i(x, \alpha_i).$$

Параметры α_i, ν_i определяются путем решения дифференциального уравнения k-го порядка

$$D \equiv \sum_{h=1}^{k} a_h f^{(h)}(x) = 0,$$

где $a_h = a_h(x, \alpha_1, \ldots, \alpha_k)$ при $h < k$ суть симметрические функции от $\alpha_1, \ldots, \alpha_k$, а $a_k = 1$. Конкретно рассмотрены: а) смесь двух экспоненциальных распределений (с параметрами соответственно α_1 и α_2) и нормального с математическим ожиданием, равным α_3 и с $\sigma^2 = 1$; б) смесь двух распределений Парето с плотностями $g_i(x) = \alpha_i x_0^{\alpha_i} x^{-1-\alpha_i}$ при $x \geqslant x_0$, $g_i(x) = 0$ при $x < x_0$, $i = 1,2$. Упомянем еще две работы [57, 74]; в них для разделения смеси двух нормальных совокупностей используются преобразования Фурье. В [10] выводятся нижние границы для дисперсии ошибки при оценивании параметров распределений, входящих в смесь; для этого используется информационное неравенство Рао—Крамера. В [44] выписаны нижняя граница Рао—Крамера для минимальной дисперсии несмещенной оценки веса ν для выборки из смеси двух распределений. Основной качественный результат этого исследования заключается в том, что даже в очень простой ситуации, такой как смесь двух экспоненциальных или смесь двух нормальных распределений со всеми известными параметрами (кроме ν), ожидаемая точность оценки ν очень низка. Приведены сжатые таблицы, позволяющие определить выборочный объем, необходимый для решения задачи разделения смеси. В [21] выведены необходимые и достаточные условия для достижения границы Рао—Крамера. Более широкие границы, использующие матрицу Бхаттачария, даны в [86].

Во всех приведенных работах строится некоторый алгоритм для нахождения оценок параметров смеси. Аналитические выражения для оценок находятся только в том случае, если вводятся некоторые предположения, еще более упрощающие задачу. Свойства полученных оценок исследованы не полностью и требуют дальнейшего изучения. Во всех задачах число возможных распределений, входящих в смесь, фиксировано, и не имеется рекомендаций для решения задачи при отсутствии такой информации. В работе [66] наряду с оцениванием параметров смеси двух нормальных распределений, основное внимание обращено на построение решающих правил для разнесения элементов выборки на две группы. Эти решающие правила охарактеризованы ожидаемым числом неправильных классификаций, и в частных случаях даны рекомендации для предпочтительного использования одного из правил.

Использование элементов выборки на протяжении всей вычислительной процедуры. Методы, описанные выше, используют элементы выборки лишь на первом этапе расчетов, а именно, для оценивания параметров смешанной совокупности; все дальнейшие вычисления производятся только с полученными статистиками. С другой стороны, имеются алгоритмы разделения смесей, в которых элементы выборки используются на протяжении всей вычислительной процедуры. Если мы при этом не располагаем заданием искомых составляющих совокупностей в виде определенных параметрических семейств (с неизвестными параметрами), то соответствующие методы принято называть методами кластер-анализа. Однако чаще всего целью кластер-анализа является решение таких задач, которые состоят в объединении элементов, являющихся самостоятельными объектами изучения, в более или менее однородные группы. В нашем обзоре группы, получаемые в результате разбиения совокупности объектов, будут рассматриваться как выборки, позволяющие судить о свойствах составляющих совокупностей (например, оценить их параметры). Поэтому в дальнейшем будем считать, что каждый элемент совокупности, подлежащей разделению, не представляет самостоятельного интереса, а важен лишь для описания составляющей совокупности, в которую он входит. Об этом аспекте применения методов кластер-анализа говорится в [55]. Отметим также, что именно смеси распределений (обычно многомерных нормальных) используются чаще всего в качестве модели для проверки эффективности конкретных методов кластер-анализа [36, 38, 39].

Полная систематизация постановок задач кластер-анализа, формулировка общей экстремальной задачи кластер-анализа в терминах достаточно общих функционалов меры внутриклассового и межклассового рассеяния дана в книге С. А. Айвазяна, З. И. Бежаевой, О. В. Староверова [1]*). Сравнение некоторых методов такого типа проводится в [65], где предложен способ оценки качества того или иного метода.

Использование мер близости. Применение любого метода группировки объектов требует введения критерия, в соответствии с которым разделение на подсовокупности будет оптимальным [11]. Чаще всего таким критерием считают условие максимума или минимума какой-либо функции от задаваемой тем или иным образом взаимной «близости» между элементами смеси.

Существует три основных способа количественного выражения близости (сходства) элементов — с помощью коэффициентов подобия, коэффициентов корреляции и различных метрических показателей расстояния (в том числе с использованием потенциальных функций). Отметим, что коэффициенты подобия

*) Обзоры работ, посвященных кластер-анализу, сделаны также в статьях: Н. Н. Райская, А. Т. Терехин, А. А. Френкель., Кластерный анализ и его применения. Заводск. лаборатория, 1972, 38, № 10, 1222—1228 (РЖМат, 1973, 2В162); Cormack R. M., A review of classification. J. Roy. Statist. Soc., 1971, A134, № 3, 321—367

используются в основном при решении задач иерархической классификации (построения так называемых дендрограмм), а коэффициенты корреляции — в задачах факторного анализа.

Выбор меры близости между элементами в значительной степени произволен, однако он некоторым образом связан с выбором критерия разделения. Последний задает условие, по которому будет оцениваться группирование элементов заданной совокупности. Правильное разбиение будет считаться выполненным, если достигнуто экстремальное значение критерия. Предполагается, что существует единственное разбиение, которое доставляет экстремум выбранному критерию, и результирующее разбиение должно совпадать с объективно существующим, если границы последнего хорошо определены. При этом критерий должен быть таким, чтобы было возможно построение эффективного алгоритма для нахождения оптимального разбиения.

Вопросы, относящиеся к выбору и использованию как мер близости, так и надлежащих критериев качества разделения, рассмотрены обстоятельно в монографии С. А. Айвазяна, З. И. Бежаевой и О. В. Староверова [1]. Там же дано подробное описание соответствующих вычислительных процедур. Учитывая также, что в этой книге приведена обширная библиография по указанным вопросам, мы считаем возможным ограничиться приведенной выше общей характеристикой данной группы методов, и за всеми дальнейшими справками отсылаем читателя к цитированной монографии [1].

Группирующие функции. В отличие от упомянутых выше методов, использующих определенные вариационные критерии, в ряде работ строятся те или иные группирующие функции для отыскания локальных максимумов плотности распределения смеси.

В [39] для элементов совокупности строится группирующая функция на основе введенного расстояния $d(x_i, x)$. Например, она полагается равной для каждой точки x числу точек, попавших в область $\Gamma=\{x: d(x_i, x)<z\}$, где z — некоторый заданный порог. Ищутся все локальные максимумы этой функции, и вся совокупность классифицируется в унимодальные группы. Доказана теорема, дающая достаточные условия, для отыскания всех локальных максимумов. Проведенные эксперименты показали независимость ошибок при разделении от типа распределения элементов в подсовокупностях. Время вычислений на IBM 360/75 для 10-мерного пространства и 1000 наблюдений находится в пределах от 3 до 6 мин. Авторы указывают на возможность расширения алгоритма для очень большого числа данных (скажем, больше 30 000). В [69] группирующая функция строится следующим образом. Каждая p-мерная точка совокупности x_k $(k=1, \ldots, N)$ рассматривается как центр p-мерного нормального распределения

$$N(x_k, \sigma) = (\sigma \cdot \sqrt{2\pi})^{-p} \exp\left\{-\frac{1}{2}(x-x_k)^T (x-x_k)\right\},$$

и составляется функция $F(x) = \sum_{k=1}^{N} N(x_k, \sigma)$. В центрах групп $F(x)$ имеет максимумы. Поэтому, выходя из каждой x_k, надо двигаться по градиенту $F(x)$ до достижения максимума, x_k, приводящие к одному и тому же максимуму $F(x)$, принадлежат к одной группе. При очень малом σ число максимумов $F(x)$ равно N, при очень больших σ число максимумов равно 1. Наметив ряд значений σ, проводят всю процедуру при каждом σ и определяют число возникающих максимумов (l). Истинно то l, которое сохраняется в сравнительно широком интервале значений σ; соответствующие l максимумов функции $F(x)$ и определяют искомые центры групп. Эта интересная работа предлагается автором для применения в случае размерности пространства не большей 10. Примеров конкретных реализаций метода не приведено. Аналогичная идея развита В. И. Васильевым, В. В. Коноваленко [2]. Там предложены следующие два типа группирующих функций:

1) оценка по методу нормальных вкладов

$$P_n(x_0) = \frac{1}{n} \sum P_n^*(x_0 - x_i),$$

где $P_n^*(x)$ — нормальная плотность, зависящая от x_i как от параметра со среднеквадратичным отклонением

$$\sigma_n = \frac{x_{(k)} - x_{(j)}}{2\sqrt{2}},$$

где $x_{(j)} < x_0 < x_{(k)}$, $x_{(j)}$ и $x_{(k)}$ — порядковые статистики;

2) оценка по методу ближайшего наблюдения

$$P_n(x_0) = \frac{p \cdot \Gamma(p/2)}{2^{p/2} \cdot n \cdot r_n^p},$$

где p — размерность выборочного пространства, r_n — радиус гиперсферы, равный среднему от суммы значений меры близости от точки x_0 до первого и второго ближайших к x_0 наблюдений смешанной совокупности. Когда значение $P_n(x_0)$ вычислено, организуется поиск ближайшей моды этой функции.

Исследованы статистические свойства оценки по методу нормальных вкладов и ближайшего наблюдения и показано, что они являются состоятельными оценками $P_n(x_0)$ в любой точке выборочного пространства. Авторы отмечают громоздкость вычислений, но эти алгоритмы имеют то преимущество, что не требуют задания извне числа групп. Предложенный подход был применен к решению задачи диагностики заболевания «инфаркт миокарда». Состояние каждого больного описывалось 240-мерным вектором (в машинных символах). Программа, реализующая алгоритм, распределила больных по 19 диагнозам. Средняя точность распознания, оцениваемая числом совпавших диагнозов машины и врача (по отношению к общему числу больных проверочной выборки), равна 75%. На ЭВМ БЭСМ-6 обработка выборки из 200 больных потребовала 5 часов машинного времени.

Дискриминантные функции. Следующие две работы [56, 84] используют идеи дискриминантного анализа для нахождения оптимального разбиения заданной совокупности элементов на однородные группы. В первой из них рассматривается частная задача, когда в наборе признаков может быть выделен один доминирующий. Сначала выполняют произвольное разбиение совокупности на две группы по этому признаку, относя в одну группу первые k элементов, ранжированных по этому признаку, и в другую группу $N—K$ оставшихся элементов. Затем полученные группы используются для построения дискриминантной функции по всем признакам, и с помощью этой функции производится перераспределение элементов по группам. Всю процедуру повторяют с разными K и выбирают то значение K^*, при котором потребуется перераспределение наименьшего числа n_{K^*} элементов. Рассмотрена дополнительная процедура на случай, когда имеется несколько значений K^*, приводящих к одному и тому же минимуму n_{K^*}. В работе не указано, в каком случае можно считать, что один из признаков является настолько доминирующим, чтобы можно было применять описанный метод. Кроме того, не доказана сходимость процедуры, и неясно, будет ли полученное разделение действительно оптимальным в общепринятом смысле. Приведенный пример показывает, что процедура хорошо работает при больших выборках ($n=100$); указывается, что она требует мало машинного времени для вычислений.

В работе [84] все признаки рассматриваются как равноправные. Здесь сначала смешанную совокупность разделяют на части при помощи линейной дискриминантной функции с произвольными (например, равными единице) коэффициентами. Затем вычисляют параметры в полученных двух частях и по ним строят новую линейную дискриминантную функцию, причем показано, что новое разделение ближе к истинному, чем предыдущее, так что итерационная процедура, построенная таким образом, сходится к оптимальному разделению в классе гиперплоскостей. Если распределения в компонентах нормальны с одинаковыми ковариационными матрицами, то сходимость процедуры можно значительно ускорить, просто экстраполируя найденный первый поворот разделяющей гиперплоскости. Если указанные условия не выполняются, то таким способом получается частный оптимум, который может служить исходным пунктом для следующей итерации. Проведя дискриминантную коррекцию, получают новое направление экстраполяции и т. д.

Для контроля процесса улучшения разделения используются третий и четвертый моменты одномерного разделения, получающегося при проектировании многомерного распределения на нормаль к разделяющей гиперплоскости. Разделение на несколько составляющих совокупностей производится методом последовательных дихотомий. Обсуждаются вопросы, связанные с использованием выборок. Несколько позже аналогичная идея была высказана в [5]. Упрощенный подход, использующий евклидову метрику, вместо метрики Махаланобиса, был предложен в [61].

Поскольку методы, описанные выше, используют выборки, то после нахождения оптимального разделения данной совокупности на группы возникает вопрос статистической значимости полученного разбиения. Этот вопрос рассматривается лишь в [34, 69, 84]. В последней работе эта значимость оценивается по известному в дискриминантном анализе критерию

$$F=\frac{N-p-1}{p}\,\frac{N}{N-2}D^2,$$

где D^2 — есть выборочная оценка расстояния Махаланобиса, а F сравнивается с критическим значением F-распределения для p и $N-p-1$ степеней свободы. В [69] полученные группы рассматриваются как нормальные и применяются известные критерии значимости.

В [34] дана таблица значений максимального расстояния между кластерами, но для одномерного случая и при нормальном распределении элементов, составляющих группы.

В работах [69] и [2] вопрос о числе групп, составляющих заданную совокупность, решается в ходе самой процедуры. В других работах при большем, чем два и неизвестном числе групп предлагается свести процесс разбиения к последовательной дихотомии [84]. Многократно применяя процесс дихотомии, можно получить разбиение на любое число групп. Однако алгоритм последовательной дихотомии сам по себе не обеспечивает оптимума при разбиении на M классов. Поэтому следует оптимизировать процесс, введя некоторый критерий объединения. Это делается в [4, 55]. Этот вопрос смыкается с вопросом задания правила остановки при иерархической группировке, так как задавая какой-то критерий остановки, мы получаем разбиение заданной совокупности на определенное число групп. В [36, 70] для определенного числа классов предлагается использовать критерий минимизации локальной энтропии. В общем случае вопрос об определении числа групп M пока не решен.

БИБЛИОГРАФИЯ

1. **Айвазян С. А.**, Бежаева З. И., Староверов О. В., Классификация многомерных наблюдений. М., Статистика, 1974. 240 с. (РЖМат, 1974, 7В424К)
2. **Васильев В. И.**, Коноваленко В. В., Непараметрический метод самообучения в распознавании образов. Распознавание образов (материалы конференции). Тарту, 1972
3. **Галушкин А. И.**, Каймин В. А., Моментный подход к решению задачи самообучения системы распознавания образов. Тр. Моск. ин-та электрон. машиностр., 1971, вып. 14, 139—146 (РЖМат, 1971, 7В411)
4. **Дорофеюк А. А.**, Алгоритмы обучения машины распознаванию образов без учителя, основанные на методе потенциальных функций. Автоматика и телемеханика, 1966, № 10, 78—87 (РЖМат, 1967, 5В423)
5. **Котюков В. И.**, О некоторых задачах таксономии (группировки) объектов. В сб. «Вычисл. системы». Вып. 50. Новосибирск, 1972, 126—135 (РЖМат, 1973, 4В538)
6. **Мальцева Н. И.**, Оценка параметров смеси мер по методу моментов. Сб. аспирантск. работ. Казанск. ун-т. Точн. науки Мат. Мех. Физ., Казань, 1969, 66—71 (РЖМат, 1970, 2В152)
7. **Миленький А. В.**, Определение статистических характеристик распознаваемых образов в режиме самообучения. Кибернетика, 1967, № 3, 56—61 (РЖМат, 1968, 6В582)
8. **Урбах В. Ю.**, К вопросу о разложении отклоняющихся от нормального статистических распределений на два нормальных распределения. I. Биофизика, 1961, 6, № 1, 3—8
9. —, К вопросу о разложении отклоняющихся от нормального статистических распределений на два нормальных распределения. II. Биофизика, 1961, 6, № 3, 266—271
10. **Хазен Э. М.**, Определение потенциальной точности решения некоторых задач распознавания и оценивания. Изв. АН СССР. Техн. кибернетика, 1971, № 3, 184—192 (РЖМат, 1971, 10В320)
11. **Цыпкин Я. З.**, Кельманс Г. К., Рекуррентные алгоритмы самообучения. Изв. АН СССР. Техн. кибернетика, 1967, № 5, 78—87 (РЖМат, 1968, 7В534)
12. **Шлезингер М. И.**, Взаимосвязь обучения и самообучения в распознавании образов. Кибернетика, 1968, № 2, 81—88
13. **Agard J.**, Melange de deux populations normales et étude de quelques fonctions $f(x, y)$ de variables normales x, y. Rev. statist. appl., 1961, **9**, № 4, 53—70 (РЖМат, 1962, 10В61)
14. **Bartlett M. S.**, Macdonald P. D. M., «Least-squares» estimation of distribution mixtures. Nature (Engl.), 1968, **217**, № 5124, 195—196 (РЖМат, 1968, 11В126)
15. **Behboodian J.**, Information for estimating the parameters in mixtures of exponential and normal distributions. Doct. diss. Univ. Mich., 1964, 122 pp. Dissert. Abstrs, 1965, **25**, № 12, part 1, 7286 (РЖМат, 1966, 7В70Д)
16. —, On a mixture of normal distributions. Biometrika, 1970, **57**, № 1, 215—217 (РЖМат, 1970, 10В102)
17. —, On the distribution of a symmetric statistic from a mixed population. Technometrics, 1972, **14**, № 4, 919—923 (РЖМат, 1973, 5В158)
18. **Bhattacharya C. G.**, A simple method of resolution of a distribution into Gaussian components. Biometrics, 1967, **23**, № 1, 115—135 (РЖМат, 1968, 8В93)
19. **Blischke W. R.**, Moment estimators for the parameters of a mixture of two binomial distributions. Ann. Math. Stat., 1962, **33**, № 2, 444—454 (РЖМат, 1963, 5В151)
20. —, Estimating the parameters of mixtures of binomial distributions.

J. Amer. Statist. Assoc., 1964, **59**, № 306, 510—528 (РЖМат, 1965, 9В66)

21. **Boes D. C.**, On the estimation of mixing distributions. Ann. Math. Stat., 1966, **37**, № 1, 177—188 (РЖМат, 1966, 9В70)
22. —, Minimax unbiased estimator of mixing distribution for finite mixtures. Sankhya. Indian J. Statist., 1967, **A29**, № 4, 417—420 (РЖМат, 1968, 12В153)
23. **Bowman K. O.**, Shenton L. R., Space of solutions for a normal mixture. Biometrika, 1973, **60**, № 3, 629—636 (РЖМат, 1974, 5В158)
24. **Bürrau C.**, The half-invariants of the sum of two typical laws of errors with an application to the problem of dissecting a frequency curve into components. Skand. aktuarietidskr., 1934, **17**, № 1, 1—5
25. **Chatterjee S. K.**, Rank approach to the multivariate two-population mixture problem. J. Multivar. Anal., 1972, **2**, № 3, 261—281 (РЖМат, 1973, 2В123)
26. **Choi K.**, Estimators for the parameters of a finite mixture of distributions. Ann. Inst. Statist. Math., 1969, **21**, № 1, 107—116 (РЖМат, 1970, 2В155)
27. —, Bulgren W. G., An estimation procedure for mixtures of distributions. J. Roy. Statist. Soc., 1968, **B30**, № 3, 444—460 (РЖМат, 1969, 11В133)
28. **Cohen A. C.**, Estimation in mixtures of discrete distributions. Proc. Int Symp. Classical and Contagious Discrete Distrib., Montreal, 1963, 373—378
29. —, Estimation in mixtures of two normal distributions. Technometrics, 1967, **9**, № 1, 15—28 (РЖМат, 1967, 11В75)
30. **Cooper D. B.**, Schwarz R. J., On suitable conditions for statistical pattern recognition without supervision. SIAM J. Appl. Math., 1969, **17**, № 5, 872—896 (РЖМат, 1970, 7В201)
31. **Day N. E.**, Estimating the components of mixture of normal distributions. Biometrika, 1969, **56**, № 3, 463—474 (РЖМат, 1970, 7В116)
32. **Deely J. J.**, Kruse R. L., Construction of sequences estimating the mixing distribution. Ann. Math. Stat., 1968, **39**, № 1, 286—288 (РЖМат, 1971, 10В212)
33. **Dick N. P.**, Bowden D. C., Maximum likelyhood estimation for mixtures of two normal distributions. Biometrics, 1973, **29**, № 4, 781—790
34. **Engelman L.**, Hartigan J. A., Percentage points of a test for clusters. J. Amer. Statist. Assoc., 1969, **64**, № 323, 1647—1648 (РЖМат, 1970, 11В130)
35. **Falls L. W.**, Estimation of parameters in compound Weibull distributions. Technometrics, 1970, **12**, № 2, 399—407 (РЖМат, 1971, 1В125)
36. **Friedman H. P.**, Rubin J., On some invariant criteria for grouping data. J. Amer. Statist. Assoc., 1967, **62**, № 320, 1159—1178 (РЖМат, 1968, 11В139)
37. **Fryer J. G.**, Robertson C. A., A comparison of some methods for estimating mixed normal distributions. Biometrika, 1972, **59**, № 3, 639—648 (РЖМат, 1973, 5В177)
38. **Fukunaga K.**, Koontz W. L. G., A criterion and an algorithm for grouping data. IEEE Trans. Comput., 1970, **19**, № 10, 917—923 (РЖМат, 1971, 8В704)
39. **Gitman I.**, Levine M. D., An algorithm for detecting unimodal fuzzy sets and its application as a clustering technique. IEEE Trans. Comput., 1970, **19**, № 7, 583—593 (РЖМат, 1971, 3В578)
40. **Gridgeman N. T.**, A comparison of two methods of analysis of mixtures of normal distributions. Technometrics, 1970, **12**, № 4, 823—833 (РЖМат, 1971, 5В149)
41. **Hasselblad V.**, Estimation of parameters for a mixture of normal distributions. Technometrics, 1966, **8**, № 3, 431—446 (РЖМат, 1967, 4В85)
42. —, Estimation of finite mixtures of distributions from the exponential

family, J. Amer. Statist. Assoc., 1969, **64**, № 328, 1459—1471 (РЖМат, 1970, 9В126)

43. **Hawkins R. H.**, A note on multiple solutions to the mixed distribution problem. Technometrics, 1972, **14**, № 4, 973—976 (РЖМат, 1973, 5В176)
44. **Hill B. M.**, Information for estimating the proportions in mixtures of exponential and hormal distributions. J. Amer. Statist. Assoc., 1963, **58**, № 304, 918—932 (РЖМат, 1964, 9В46)
45. **Hosmer D. W.**, On MLE the parameters of a mixture of two normal distributions when the sample size is small. Communic. Stat., 1973, **1**, № 3, 217—225
46. **Hosmer David W., Jr.** A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. Biometrics, 1973, **29**, № 4, 761—770 (РЖМат, 1974, 6В132)
47. **Joffe A. D.**, Mixed exponential estimation by the method of half moments. Appl. Statist., 1964, **13**, № 2, 91—98 (РЖМат, 1966, 9В108)
48. **John E.**, Bayesian estimation of mixture distributions. Ann. Math. Stat., 1968, **39**, № 4, 1289—1302
49. **John S.**, On identifying the population of origin of each observation in a mixture of observations from two normal populations. Technometrics, 1970, **12**, № 3, 553—563 (РЖМат, 1971, 2В127)
50. —, On identifying the population of origin of each observation in a mixture of observations from two gamma populations. Technometrics, 1970, **12**, № 3, 565—568 (РЖМат, 1971, 3В96)
51. **Kabir A. B. M.**, Estimation of parameters of a finite mixture of distributions. J. Roy. Statist. Soc., 1968, **B30**, № 3, 472—482 (РЖМат, 1969, 10В86)
52. **Kaski E.**, Krysicki W., Die Parameter schätzung einer Mischung von zwei Laplacischen Verteilungen (in allgemeinem Fall). Prace Mat., 1967, **11**, ser. 1, 23—31
53. **Krysicki W.**, Zastosowanie metody momentów do estymacji parametrów mieszaniny dwóch rozkładów Laplace'a. Zesz. nauk. Politechn. lódzk., 1966, № 77, 5—13 (РЖМат, 1966, 11В75)
54. —, Estimation of the parameters of the mixture of an arbitrary number of exponential distributions. Demonstr. math., 1972, **4**, № 3, 175—183 (РЖМат, 1973, 5В175)
55. **Mac Queen J.**, Some methods for classification and analysis of multivariate observations. Proc. 5th Berkeley Sympos. Math. Statist. and Probabil., 1965—1966. Vol. 1. Berkeley—Los Angeles, 1967, 281—297 (РЖМат, 1970, 3В168)
56. **Mayer L. S.**, A method of cluster analysis when there exist multiple indicators of a theoretic concept Biometrics, 1971, **27**, № 1, 143—155
57. **Medgyessy P.**, Varga L., Gauss-függvény keverékek numerikus felbontására szolgáló egyik eljárás javításáról. 1—4. Magy. tud. akad. Mat. és fiz. tud. oszt. közl., 1968, **18**, № 1, 31—39 (РЖМат, 1969, 11В118)
58. **Meeden G.**, Bayes estimation of the mixing distribution, the discrete case. Ann. Math. Stat., 1972, **43**, № 6, 1993—1999 (РЖМат, 1973, 7В148)
59. **Mohanty N. C.**, On the identifiability of finite mixtures of Laguerre distributions. IEEE Trans. Inform. Theory, 1972, **18**, № 4, 514—515 (РЖМат, 1973, 1В209)
60. **Molenaar W.**, Survey of estimation methods a mixture of two normal distributions. Statist. neer., 1965, **19**, № 4, 249—265
61. **Nagy G.**, Shelton G. L., Self-corrective character recognition system. IEEE Trans. Inform. Theory, 1966, **12**, № 2, 215—222 (РЖМат, 1967, 3В438)
62. **Patrick E. A.**, Costello J. P., On unsupervised estimation algorithms. IEEE Trans. Inform. Theory, 1970, **16**, № 5, 556—569 (РЖМат, 1971, 5В225)
63. **Pearson K.**, Contribution to the mathematical theory of evolution. Philos.

Trans. Roy. Soc. London, 1894, A185, 71—110 (Karl Pearson's early statistical papers, Cambridge, 1948, 1—40)
64. **Preston P. F.**, Estimating the mixing distribution by piece-wise polynomial arcs. Austral. J. Statist., 1971, 13, № 2, 64—76 (РЖМат, 1972, 4В123)
65. **Rand W. M.**, Objective criteria for the evaluation of clustering methods. J. Amer. Statist. Assoc., 1971, 66, № 336, 846—850 (РЖМат, 1972, 8В183)
66. **Rayment P. R.**, The identification problem for a mixture of observations from two normal populations. Technometrics, 1972, 14, № 4, 911—918 (РЖМат, 1973, 5В188)
67. **Rider P. R.**, The method of moments applied to a mixture of two exponential distributions. Ann. Math. Stat., 1961, 32, № 1, 143—147 (РЖМат, 1962, 1В66)
68. —, Estimating the parameters of mixed Poisson, binomial and Weibull distributions. Bull. Int. Statist. Inst., 1962, 39, № 11, 225—232
69. **Schnell P.**, Eine Methode zur Auffindung von gruppen. Biometr. Z., 1964, 6, № 1, 47—48
70. **Scott A. J.**, Symons M. J., Clustering methods based on likelihood ratio criteria. Biometrics, 1971, 27, № 2, 387—397 (РЖМат, 1971, 12В346)
71. **Singh M. P.**, A note on generalized inflated binomial distribution. Sankhya. Indian J. Statist., 1966, A28, № 1, 99 (РЖМат, 1967, 7В82)
72. **Sleeba J.**, On analyzing mixed samples. J. Amer. Statist. Assoc., 1970, 65, № 330, 755—762 (РЖМат, 1971, 2В135)
73. **Sokal R. R.**, Sneath P. H. A., Principles of numerical taxonomy. London, Freeman, 1963
74. **Stanat D. F.**, Nonsupervised pattern recognition through the decomposition of probability functions. Doct. diss. Univ. Mich., 1966, 60 pp. Dissert. Abstrs., 1967, B27, № 7, 2452 (РЖМат, 1968, 4В116Д)
75. **Strömgren B.**, Tables and diagrams for dissecting a frequency curve into components by the halfinvariant method. Skand. aktuarietidskr., 1934, 17, № 1, 7—54
76. **Tallis G. M.**, Light R., The use of fractional moments for estimating the parameters of a mixed exponential distribution. Technometrics, 1968, 10, № 1, 161—175 (РЖМат, 1968, 11В129)
77. **Tan W. Y.**, Chang W. C., Come comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of normal densities. Biometrics, 1971, 27, № 2, 489
78. —, —, Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of two normal densities. J. Amer. Statist. Assoc., 1972, 67, № 339, 702—708 (РЖМат, 1973, 4В178)
79. **Teicher H.**, On the mixture of distributions. Ann. Math. Stat., 1960, 31, № 1, 55—73 (РЖМат, 1962, 1В21)
80. —, Identifiability of mixtures. Ann. Math. Stat., 1961, 32, № 1, 244—248 (РЖМат, 1964, 1В9)
81. —, Identifiability of finite mixtures. Ann. Math. Stat., 1963, 34, № 4, 1265—1269 (РЖМат, 1965, 10В6)
82. —, Identifiability of mixtures of product measures. Ann. Math. Stat., 1967, 38, № 4, 1300—1302 (РЖМат, 1971, 10В206)
83. **Thionet P.**, Note sur les mélanges de certaines distributions de probabilités. Publs. Inst. statist. Univ. Paris, 1966, 15, № 1, 61—80 (РЖМат, 1967, 6В62)
84. **Urbakh V. Yu.**, A discriminant method of clustering. J. Multivar. Anal., 1972, 2, № 3, 249—260 (РЖМат, 1973, 1В287)
85. **Wasilewski M. J.**, Über gewisse Merkmale der Mischung von zwei Maxwellschen — Verteilungen. Zesz. nauk. Politechn. łódzk., 1967, № 95, 5—21 (РЖМат, 1969, 2В97)
86. **Whittaker J.**, The Bhattacharyya matrix for the mixture of two distributions. Biometrika, 1973, 60, № 1, 201—202 (РЖМат, 1973, 9В75)

87. **Wolfe J. H.**, Pattern clustering by multivariate mixture analysis. Multivariate Behavior. Res., 1970, 5, 329
88. **Yakowitz S. J.**, A consistent estimator for the identification of finite mixtures. Ann. Math. Stat., 1969, 40, № 5, 1728—1735 (РЖМат, 1971, 7В295)
89. —, Unsupervised learning and the identification of finite mixtures. IEEE Trans. Inform. Theory, 1970, 16, № 3, 330—338 (РЖМат, 1971, 2В226)
90. —, Spragins J. D., On the identifiability of finite mixtures. Ann. Math. Stat., 1968, 39, № 1, 209—214 (РЖМат, 1971, 9В118)
91. **Young T. Y.**, Coraluppi G., Stochastic estimation of a mixture of normal density functions using an information criterion. IEEE Trans. Inform. Theory, 1970, 16, № 3, 258—263 (РЖМат, 1971, 1В118)