

Math-Net.Ru

Общероссийский математический портал

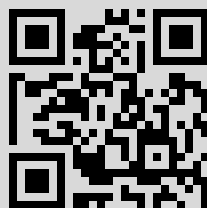
А. М. Шурыгин, Размерности многомерной статистики, *Автомат. и телемех.*, 1995, выпуск 8, 103–123

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 85.235.162.70

18 сентября 2018 г., 11:38:20



РАЗМЕРНОСТИ МНОГОМЕРНОЙ СТАТИСТИКИ

Рассматривается устойчивость оценок минимума контраста параметров многомерного нормального распределения. Вариационная производная нормированной дисперсии оценки по плотности распределения рассматривается как мера неустойчивости оценки к малым отклонениям распределения от модельного. Через нее определяется устойчивость оценки, меняющаяся от нуля для оценки максимального правдоподобия до единицы для оценки максимальной устойчивости. Вариационная производная оценки максимальной устойчивости растет как показательная функция размерности пространства, что исключает возможность получения устойчивых оценок для большого числа зависимых признаков. Этого недостатка лишено предлагаемое многомерное сферическое распределение, его одномерные проекции по распределениям близки к нормальным.

Основной моделью многомерной статистики является совместное нормальное распределение зависимых признаков. На этой модели решены многие задачи, полезные в приложениях. В учебниках и руководствах не указывается возможное количество таких признаков. В пакетах программ существует ограничение: количество признаков должно быть меньше количества наблюдений (иначе матрица ковариаций окажется вырожденной). Но на многих реальных задачах выяснилось, что и при соблюдении этого условия решение по небольшому количеству признаков оказывается лучшим, чем по полному их набору. Решение улучшается также, если, игнорируя слабые корреляции, разделить множество признаков на псевдонезависимые подмножества, решить задачу для каждого из получившихся маломерных распределений, а затем объединить эти частные решения в общее. Это наводит на мысль, что в приложениях многомерной статистики существует сильное ограничение количества зависимых признаков.

1. Дискриминантный анализ и распознавание образов

В 60-х годах в связи с развитием вычислительной техники появилось много программ, которые по множеству признаков решали задачу классификации. В простейшем случае по образцам из двух классов вырабатывалось правило, дававшее возможность определить принадлежность нового объекта к одному из этих классов. В идеологии решения этой задачи сразу наметились два направления. Первое, статистическое, исходило из предположения о многомерной нормальности распределения каждого из классов; эффективность получаемого правила классификации проверялось моделированием многомерных нормальных совокупностей. Во втором направлении информация кодировалась так, что можно было использовать алгоритмы распознавания зрительных образов; объекты с известной принадлежностью делились на "обучение" и "экзамен": по "обучению" вырабатывалось правило классификации, на "экзамене" проверялась его эффективность. Два направления в решении одной задачи конкурировали, по крайней мере хорошие "распознаватели" сравнивали свое решение со статистической дискриминантной функцией и на "экзамене" показывали, что их решение лучше статистического, специально полученного для решения этой задачи. Бонгард, создатель популярного алгоритма "Кора" [1],

выступая на конференции в Московском нефтяном институте, говорил примерно следующее: "Я делал алгоритм, который должен был различать зрительные образы. Алгоритм оказался неудачным: он не может отличить треугольник от квадрата. Я не знаю, почему мой алгоритм различает нефтеносные и водоносные пласты по набору геофизических признаков". Чтобы найти ответ, автор просмотрел первичные геофизические материалы. Оказалось, что существуют два признака, по которым нефтеносные и водоносные пласты почти полностью разделяются. Но оставалось непонятным, почему в такой простой ситуации плохо работала многомерная статистика: одномерные распределения не имели грубых отклонений от нормальности.

На другом материале автор [2] сравнил десять методов решения задачи классификации. На "экзамене" стандартные методы многомерной статистики оказались наихудшими, "Кора" показала средние результаты, а наилучшими были статистические решения по небольшому количеству специально отобранных признаков и метод [3] сведения решения к последовательности двумерных задач.

В учебниках по многомерной статистике приводятся примеры решения реальных задач, но небольшой размерности. Так, в [4] приводятся четырехмерные задачи: антропометрическая по данным [5] и биометрическая по данным [6]. А. Н. Колмогоров, предложивший асимптотику, в которой количество признаков растет пропорционально длине выборки [7], советовал ученикам делать трехмерные модели выборки, втыкая в ящик с песком спицы с шариками на концах. Если специалисты конкретных наук стараются увеличить количество признаков, характеризующих объекты, то статистики разработали серию приемов, уменьшающих размерность задачи [8]. Создается впечатление, что многомерная статистика почему-то не может решать задачи большой размерности.

Ограниченность возможностей многомерной статистики может быть понятой в связи с исследованиями устойчивости оценок. Рассмотрим эту задачу подробнее.

2. Оценки минимума контраста и их устойчивость

Пусть $x_1, \dots, x_n \in R^p$ – векторы, имеющие p -мерное нормальное распределение $N_p(m, C)$ с плотностью распределения (п.р.) $(2\pi)^{-p/2} C^{-1/2} \exp(-(x-m)^T C^{-1} (x-m)/2)$, где символ T использован для обозначения транспонирования. Оценками максимума правдоподобия (ОМП) для вектора математических ожиданий m и матрицы ковариаций C являются $\hat{m} = n^{-1} \sum x_i$ и $\hat{C} = n^{-1} \sum (x_i - m)(x_i - m)^T$. Среди возможных оценок они обладают наименьшей асимптотической дисперсией, но оказываются чрезвычайно неустойчивыми даже при малом нарушении нормальности распределения [9].

Устойчивость оценок можно повысить за счет снижения их эффективности. Так, Л. Д. Мешалкин в [8, 10] показал состоятельность оценок \hat{m} и \hat{C} из решаемой итерациями системы уравнений

$$(1) \quad \begin{aligned} \Sigma(x_i - \hat{m}) \exp(-\lambda q_i/2) &= 0, \\ \Sigma \left[(x_i - \hat{m})(x_i - \hat{m})^T - (1 + \lambda)^{-1} \hat{C} \right] \exp(-\lambda q_i/2) &= 0, \end{aligned}$$

где $q_i = (x_i - \hat{m})^T C^{-1} (x_i - \hat{m})$, а $\lambda > 0$ – произвольная константа, итерациями отыскиваются \hat{m} и \hat{C}^{-1} . В [11, 12] выписаны вторые моменты этих оценок.

В [13] предложены оценки из системы уравнений

$$(2) \quad \begin{aligned} \Sigma(x_i - \hat{m}) \xi(\sqrt{q_i}) &= 0, \\ \Sigma(x_i - \hat{m})(x_i - \hat{m})^T \xi^2(\sqrt{q_i}) &= n\lambda \hat{C}, \end{aligned}$$

где

$$\xi(\sqrt{q}) = \begin{cases} 1 & \text{при } \sqrt{q} < k, \\ k/\sqrt{q} & \text{при } \sqrt{q} > k \end{cases}$$

– весовая функция Хьюбера с величиной $k > 0$, зависящей от неопениваемой доли загрязняющих наблюдений в выборке, и выписана функция $\chi = \chi(p, k)$, обеспечивающая состоятельность оценки \hat{C} .

Авторы отмечали снижение эффективности оценок при росте размерности p и сильную их зависимость от произвольных параметров λ и k . Чтобы избавиться от этих недостатков, в [14] была предложена модель выборки, которая с вероятностью $n^{-\zeta/2}$, $0 < \zeta < 1$ имеет точечное загрязнение, распределенное случайно в серии выборок. Наилучшие оценки для наихудшего нормального распределения загрязняющего импульса являются решениями системы уравнений (1) при $\lambda = (p+1)^{-1}$ в первом уравнении системы и $\lambda = 2(p+2)^{-1}$ во втором уравнении. С ростом размерности p эффективность этих оценок увеличивается.

В теории робастности для оценки центра нормального распределения предложены различные весовые функции типа экспоненты в (1) и $\xi(\sqrt{q})$ в (2), которые занижают веса периферийных наблюдений. Весовые функции симметричны, это обеспечивает состоятельность оценки \hat{m} , но занижает оценки дисперсий. Выписывание поправочного коэффициента типа $(1+\lambda)^{-1}$ в (1) и $\chi = \chi(p, k)$ в (2), обеспечивающих состоятельность оценки \hat{C} , в общем случае сложно из-за необходимости интегрировать в R^p . Но многомерный интеграл сводится к одномерному [15].

Лемма. Пусть функция $\eta(q)$ такая, что оценка \hat{C} матрицы ковариаций C из матричного уравнения

$$\Sigma \left[(x_i - m)(x_i - m)^T - \beta \hat{C} \right] \eta(q_i) = 0$$

состоятельна. Тогда

$$\beta = \beta(p) = \frac{2}{p} \int_0^\infty s^{p/2} e^{-s} \eta(2s) ds \bigg/ \int_0^\infty s^{p/2-1} e^{-s} \eta(2s) ds.$$

Коэффициент $\beta(p)$ не зависит от p тогда и только тогда, когда весовая функция $\eta(q_i) = \exp(-\lambda q_i/2)$ экспоненциальна: $\beta = (1+\lambda)^{-1}$.

Для повышения устойчивости оценки нужно определить меру устойчивости, что и сделано в [16] при рассмотрении вариационной производной нормированной дисперсии оценки по п.р. Предложенный подход обобщается на многомерный случай автоматически следующим образом.

Пусть $f(x, \theta)$ – п.р. вектора $x \in R^p$, зависящая от параметра $\theta \in \Theta \subset R^1$, и функция $\psi(x, \theta)$ такая, что оценка минимума контраста $\hat{\theta}$ из уравнения

$$(3) \quad \Sigma \psi(x_i, \hat{\theta}) = 0$$

является состоятельной. Тогда ее нормированная асимптотическая дисперсия

$$(4) \quad V(f, \psi) = \lim n E (\hat{\theta} - \theta)^2 = E \psi^2(x) \left[E \dot{\psi}(x, \theta) \right]^{-2},$$

где $\dot{\psi}(x, \theta) = (d/d\theta) \psi(x, \theta)$. Вариационная производная этого функционала по п.р. равна

$$(5) \quad W(f, \psi) = \frac{\delta}{\delta f} V(f, \psi) = \int_{R^p} \psi^2(x) dx \left[E \dot{\psi}(x, \theta) \right]^{-2} > 0$$

и достигает минимума $W^* = W(f, \psi^*)$ на функции

$$\psi^*(x, \theta) = \arg \min_{\psi} W(f, \psi) = c \left[\frac{d}{d\theta} f(x, \theta) + \alpha f(x, \theta) \right],$$

где $c = c(\theta) \neq 0$ - произвольный множитель, а $\alpha = \alpha(\theta)$ такая, что $E \psi^*(x) = 0$; соответствующая оценка $\hat{\theta}^*$ называется оценкой максимума устойчивости (ОМУ). Величина

$$\text{stb } \hat{\theta} = W^* / W(f, \psi)$$

называется устойчивостью оценки $\hat{\theta}$ из (3). Она построена по аналогии с эффективностью оценки $\hat{\theta}$:

$$\text{eff } \hat{\theta} = \overset{\circ}{V} / V(f, \psi),$$

где $\overset{\circ}{V}$ - нормированная асимптотическая дисперсия ОМП.

Следуя терминологии, установившейся в прикладной математике и механике, оценку следует называть устойчивой, если малые изменения модельной п.р. $f(x)$ ведут к малым изменениям дисперсии $V(f, \psi)$ и неустойчивой, если эти изменения могут быть не малыми. Легко видеть, что необходимым и достаточным условием устойчивости является конечность вариационной производной $W(f, \psi)$. Это условие не выполняется для ОМП параметров нормального распределения. Например, в случае $x \in R^1$ для среднего \bar{x} с $\psi(y) = y$, $y = x - t$ интеграл $W(f, \psi) \sim \int_{-\infty}^{\infty} \psi^2(y) dy = \int_{-\infty}^{\infty} y^2 dy$ расходится; здесь знаком \sim обозначена эквивалентность (в этом отноше-

нии среднее \bar{x} хуже медианы с $\psi(y) = \text{sgn}(y)$, для которой интеграл $W(f, \psi) \sim \int_{-\infty}^{\infty} dy$ расходится медленнее). Если сделать предположение о распределении, усеченном на уровне ε^{-1} , где $\varepsilon > 0$ мало, то интеграл $W(f, y)$ будет расти, как ε^{-3} , и устойчивость среднего составит $\text{stb } \bar{x} = O(\varepsilon^3)$.

Рассмотрим вариационную производную W_p в многомерном распределении $N_p(m, I)$, где I - единичная матрица, предположив его усеченным на уровне $|x - \mu| = \varepsilon^{-1} = 3$ по каждой из координат. Тогда нетрудно получить интегрированием, что ОМП $\hat{\mu} = \bar{x}$ произвольной компоненты μ вектора m имеет $W_p \approx 3 \cdot 6^p$ (это верно не только для нормального, но и для любого многомерного распределения), так что прибавление каждого нового зависимого признака увеличивает вариационную производную примерно в шесть (!) раз. Многомерная статистика использует для оценки параметров средние величины, являющиеся оценками максимума правдоподобия для нормального распределения, и тенденция опытных статистиков уменьшать размерность задачи является естественным стремлением спастись от катастрофического падения устойчивости оценок при росте размерности.

Второй способ спасения - это замена неустойчивых оценок максимума правдоподобия устойчивыми оценками. Они есть: например, оценки Мешалкина из системы (1) устойчивы при любом $0 < \lambda < 1$ (но оценки Марроны из системы (2) не устойчивы при всех k , так как $\psi(x) \notin L_2$). Имея две характеристики оценки $\hat{\theta}$, $\text{eff } \hat{\theta}$ и $\text{stb } \hat{\theta}$, можно выбрать в некотором смысле наилучшую оценку. В [16] предложена следующая оптимизация.

При фиксированной эффективности наибольшая устойчивость (или при фиксированной устойчивости наибольшая эффективность) достигается на условно-оптимальных оценках $\tilde{\theta}$ с

$$\psi(\mathbf{x}, \theta) = c \left[\frac{d}{d\theta} \ln f(\mathbf{x}, \theta) + \alpha \right] [1 + \gamma/f(\mathbf{x})]^{-1},$$

где константа $\gamma = \gamma(\theta) > 0$ определяется выбранным уровнем эффективности (или устойчивости). Условно-оптимальная оценка $\tilde{\theta}$, минимизирующая сумму $(\text{eff } \tilde{\theta})^{-1} + (\text{stb } \tilde{\theta})^{-1}$, называется компромиссной, для нее $\gamma = \dot{V} / \dot{W}$. Компромиссная оценка может быть аппроксимирована корневой оценкой $\tilde{\theta}$ с

$$\psi(\mathbf{x}, \theta) = c \left[\frac{d}{d\theta} \ln f(\mathbf{x}, \theta) + \alpha \right] \sqrt{f(\mathbf{x})}, \quad \alpha = \alpha(\theta) : E\psi(\mathbf{x}) = 0,$$

для нее $\text{eff } \tilde{\theta} = \text{stb } \tilde{\theta}$.

Теорема 1. Для распределения $N_p(\mathbf{m}, \mathbf{C})$ условно-оптимальные оценки $\tilde{\mathbf{m}}$ и $\tilde{\mathbf{C}}$ удовлетворяют системе уравнений

$$\begin{aligned} \Sigma(\mathbf{x}_i - \tilde{\mathbf{m}}) \left[1 + \gamma_1(2\pi)^{p/2} \exp(\tilde{q}_i/2) \right]^{-1} &= 0, \\ \Sigma \left[(\mathbf{x}_i - \tilde{\mathbf{m}})(\mathbf{x}_i - \tilde{\mathbf{m}})^T - \beta \tilde{\mathbf{C}} \right] \left[1 + \gamma_2(2\pi)^{p/2} \exp(\tilde{q}_i/2) \right]^{-1} &= 0, \end{aligned}$$

где $\tilde{q}_i = (\mathbf{x}_i - \tilde{\mathbf{m}})^T \tilde{\mathbf{C}}^{-1} (\mathbf{x}_i - \tilde{\mathbf{m}})$, а

$$\beta = \beta(p) = \frac{2 \int_0^\infty s^{p/2} e^{-s} [1 + \gamma_2(2\pi)^{p/2} e^s]^{-1} ds}{p \int_0^\infty s^{p/2-1} e^{-s} [1 + \gamma_2(2\pi)^{p/2} e^s]^{-1} ds}.$$

Условно-оптимальная оценка $\tilde{\mu}$ произвольной компоненты μ вектора \mathbf{m} имеет вид

$$\begin{aligned} \text{eff } \tilde{\mu} &= \left[\int_0^\infty \frac{s^{p/2} e^{-s} ds}{1 + \gamma_1(2\pi)^{p/2} e^s} \right]^2 / \Gamma(p/2 + 1) \int_0^\infty \frac{s^{p/2} e^{-s} ds}{(1 + \gamma_1(2\pi)^{p/2} e^s)^2}, \\ \text{stb } \tilde{\mu} &= 2^{p/2+1} \left[\int_0^\infty \frac{s^{p/2} e^{-s} ds}{1 + \gamma_1(2\pi)^{p/2} e^s} \right]^2 / \Gamma(p/2 + 1) \int_0^\infty \frac{s^{p/2} ds}{(1 + \gamma_1(2\pi)^{p/2} e^s)^2}. \end{aligned}$$

Компромиссные оценки $\check{\mathbf{m}}$ и $\check{\mathbf{C}}$ удовлетворяют системе уравнений

$$\begin{aligned} \Sigma(\mathbf{x}_i - \check{\mathbf{m}}) \left[1 + \exp(\check{q}_i/2)/2^{p/2+1} \right]^{-1} &= 0, \\ \Sigma \left[(\mathbf{x}_i - \check{\mathbf{m}})(\mathbf{x}_i - \check{\mathbf{m}})^T - \beta \check{\mathbf{C}} \right] \left[1 + \exp(\check{q}_i/2)/2^{p/2+2} \right]^{-1} &= 0. \end{aligned}$$

Таблица 1

p	1	2	3	4	5	6	7	8	9	10
$\beta(p)$	0,745	0,744	0,742	0,740	0,738	0,737	0,736	0,734	0,733	0,732

Корневые оценки $\tilde{\mathbf{m}}$ и $\tilde{\mathbf{C}}$ удовлетворяют системе уравнений

$$\begin{aligned}\Sigma(\mathbf{x}_i - \tilde{\mathbf{m}}) \exp(-\tilde{q}_i/4) &= 0, \\ \Sigma \left[(\mathbf{x}_i - \tilde{\mathbf{m}})(\mathbf{x}_i - \tilde{\mathbf{m}})^T - \frac{2}{3}\tilde{\mathbf{C}} \right] \exp(-\tilde{q}_i/4) &= 0,\end{aligned}$$

и для произвольной компоненты μ вектора \mathbf{m} и диагонального элемента σ^2 матрицы \mathbf{C}

$$\begin{aligned}\text{eff } \tilde{\mu} &= \text{stb } \tilde{\mu} = (8/9)^{p/2+1}, \\ \text{eff } \tilde{\sigma}^2 &= \text{stb } \tilde{\sigma}^2 = (16/19)(8/9)^{p/2+1}.\end{aligned}$$

Значения константы $\beta(p)$ для вычисления компромиссной оценки $\tilde{\mathbf{C}}$ были получены численным интегрированием и приведены в табл. 1. Предельное значение $\beta(\infty) = \ln 2$ получено асимптотическим интегрированием методом Лапласа.

На рис. 1, а в системе координат $(\text{eff } \tilde{\mu}, \text{stb } \tilde{\mu})$ изображены траектории условно-оптимальных оценок центра нормального распределения для размерностей $p = 1, \dots, 20$. Значения p помещены в точках, отвечающих компромиссным оценкам, они расположены несколько выше биссектрисы координатного угла; значения $\eta = -\lg((2\pi)^{p/2}\gamma)$ оцифрованы по левому краю поля, занятого траекториями. При $\gamma = 0$ траектории сходятся в точке $(1, 0)$, отвечающей ОМП $\hat{\mu}$. Пересечения траекторий с верхней стороной квадрата при $\gamma = \infty$ соответствуют ОМУ $\hat{\mu}^*$. Компромиссные оценки действительно являются компромиссом между крайностями – ОМП и ОМУ. При росте размерности p уменьшается эффективность и устойчивость компромиссных оценок, при $p = 13$ примерно на половину, что указывает на неизбежно плохое качество решения реальных задач с большим количеством зависимых признаков.

Условно-оптимальные оценки сравним с оценками Мешалкина из системы (1). При $\lambda = 0$ они совпадают с ОМП, при $\lambda = 1$ с ОМУ, а при $\lambda = 1/2$ с корневыми оценками. Используя формулы (4) и (5), дифференцируя и интегрируя, найдем, что для экспоненциально взвешенной оценки $\hat{\mu}$ произвольной компоненты μ вектора \mathbf{m}

$$\begin{aligned}\text{eff } \hat{\mu} &= [(1 + 2\lambda)(1 + \lambda)^{-2}]^{p/2+1}, \\ \text{stb } \hat{\mu} &= [4\lambda(1 + \lambda)^{-2}]^{p/2+1},\end{aligned}$$

а для оценки $\hat{\sigma}^2$ произвольного диагонального элемента σ^2 матрицы \mathbf{C}

$$\begin{aligned}\text{eff } \hat{\sigma}^2 &= 2(1 + 2\lambda)^{p/2+2}(2 + 4\lambda + 3\lambda^2)^{-1}(1 + \lambda)^{-p-2}, \\ \text{stb } \hat{\sigma}^2 &= 2^{p+5}\lambda^{p/2+2}(3 + 2\lambda + 3\lambda^2)^{-1}(1 + \lambda)^{-p-2};\end{aligned}$$

значения эффективностей были выписаны в [11, 12]. На рис. 1, б изображены траектории оценок центра в координатах $(\text{eff } \hat{\mu}, \text{stb } \hat{\mu})$, значению $\lambda = 1/2$ отвечают корневые оценки, расположенные на биссектрисе координатного угла. На рис. 1, б

кружками отмечены стойкие оценки, полученные в [14] при максиминной (по эффективности) оптимизации оценивания в загрязненных выборках. При $p = 1$ стойкая и корневая оценки совпадают. При больших p высокая эффективность стойких оценок сохраняется за счет снижения устойчивости, которая не фигурировала в постановке задачи.

На рис. 1, в изображены траектории оценок Мешалкина $\hat{\sigma}^2$ для дисперсии σ^2 . Обозначенные кружками стойкие оценки при росте p сохраняют высокую эффективность за счет снижения устойчивости; при $p = 2$ стойкая и корневая оценки совпадают.

Важным выводом проведенного сравнения является близость корневых оценок к компромиссным, особенно при небольших размерностях. Если их различиями можно пренебречь, то аналитически простые корневые оценки можно использовать для аппроксимации более сложных компромиссных оценок. Покажем полезность такой аппроксимации на трех примерах.

Пример 1. Регуляризация компромиссной оценки матрицы ковариаций. Рассмотрим сначала ОМП \hat{C} . В расчетах обычно используется матрица \hat{C}^{-1} , сильно зависящая от определителя $|\hat{C}|$, который имеет смещение:

$$(6) \quad E|\hat{C}| = |C| [1 - p(p+1)/2n + O(n^{-2})].$$

Пусть, например, $p = 5$, а $n = 30$, тогда смещение $O(n^{-1})$ составляет половину оцениваемой величины. Чтобы избавиться от этого недостатка, в [17] предложено изменить член $O(n^{-1})$ в оценке \hat{C} так, чтобы член $O(n^{-1})$ в разложении $E|C|$ по степеням n^{-1} обратиться в нуль и соответствующую оценку назвать обусловленной. Было показано моделированием, что при решении дискри-

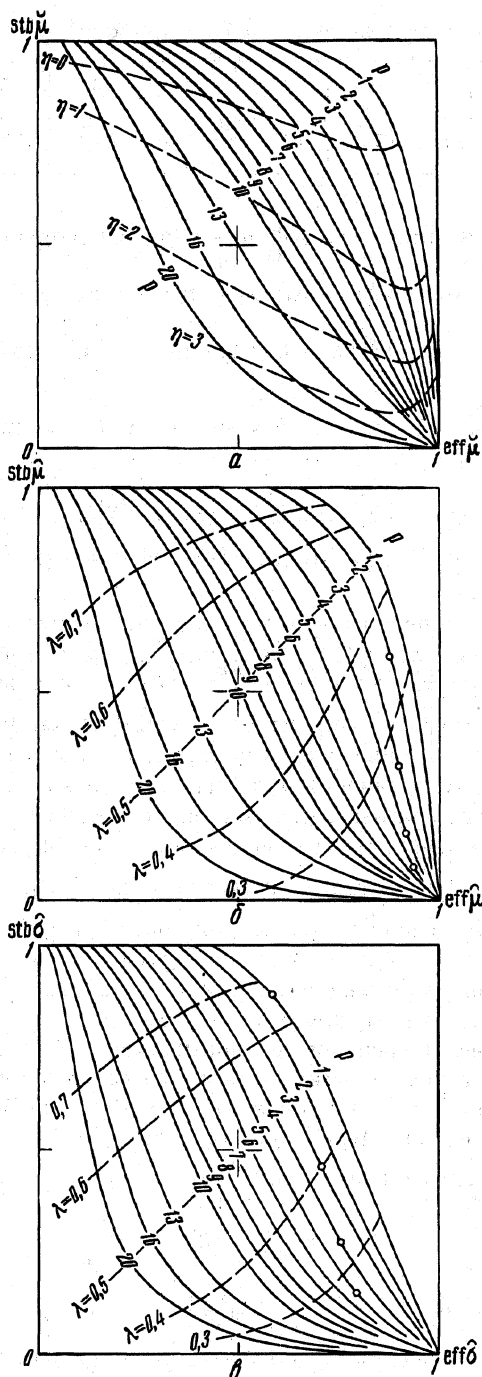


Рис. 1. Эффективности и устойчивости а) условно-оптимальных оценок $\hat{\mu}$ центра распределения, б) оценок Мешалкина $\hat{\mu}$ для центра распределения, в) оценок Мешалкина $\hat{\sigma}^2$ для дисперсии, в зависимости от количества признаков p

минантной задачи ОМП $\hat{\mathbf{C}}$ хуже обусловленной ридж-оценки, а последняя хуже обусловленной диагональной оценки вида

$$\hat{\mathbf{C}} = (1 - \nu) \mathring{\mathbf{C}} + \nu \mathring{\mathbf{C}}_{dg}, \quad \nu = O(n^{-1}) : E|\hat{\mathbf{C}}| = |\mathbf{C}|[1 + O(n^{-2})],$$

где $\mathring{\mathbf{C}}_{dg}$ – диагональ матрицы $\mathring{\mathbf{C}}$.

При росте размерности вырождается любая эквивариантная оценка, в том числе и компромиссная. Чтобы получить обусловленную компромиссную оценку, нужно громоздкое численное интегрирование, которого можно избежать, аппроксимировав компромиссную оценку корневой $\mathring{\mathbf{C}}$. Тогда для обусловленной корневой оценки

$$\hat{\mathbf{C}} = (1 - \nu) \tilde{\mathbf{C}} + \nu \tilde{\mathbf{C}}_{dg}, \quad \nu = (3/2\sqrt{2})^{p+4} p(p+1) [2n(\text{tr } \mathbf{R}^{-1} - p)]^{-1},$$

где $\mathbf{R} = \mathring{\mathbf{C}}_{dg}^{-1/2} \mathring{\mathbf{C}} \mathring{\mathbf{C}}_{dg}^{-1/2}$ – матрица корреляций, соответствующая матрице ковариаций $\mathring{\mathbf{C}}$. При получении диагональной компромиссной оценки матрицы ковариаций можно использовать значение ν , выписанное выше для обусловленной диагональной корневой оценки $\hat{\mathbf{C}}$. При итерациях в качестве начального приближения лучше брать $\mathring{\mathbf{C}}_{dg}$, чтобы обратная матрица существовала и при вырожденном распределении.

Пример 2. Критерий "однородности" выборки. В предположении о смеси различных распределений алгоритмы кластер-анализа [18] делят наблюдения на сравнительно компактные группы (кластеры). Слабым местом в решении этой задачи обычно является наличие константы, определяющей контрастность деления: при одном ее крайнем значении все наблюдения оказываются в одном кластере, при другом крайнем значении каждое наблюдение оказывается отдельным кластером.

Проверка гипотезы о том, что наблюдения могут рассматриваться как выборка из одного нормального распределения, может быть проведена при помощи компромиссных или корневых оценок. В смеси распределений эти взвешенные оценки при итерациях сходятся к параметрам распределения группы, наибольшей по количеству наблюдений, так что наблюдения из других групп оказываются на далекой периферии этого распределения и величина

$$K = (np)^{-1} \Sigma (\mathbf{x}_i - \tilde{\mathbf{m}})^T \tilde{\mathbf{C}}^{-1} (\mathbf{x}_i - \tilde{\mathbf{m}})$$

оказывается аномально большой [11, 19]. При "однородности"

$$\sqrt{np} (K - 1) \rightarrow N(0, \xi^2).$$

$$\xi^2 = \frac{19}{16} \frac{3^{p+2}}{2^{3p/2}} + \frac{1}{4} (p-1)^2 \frac{3^{p+2}}{2^{p/2+2}} - 2.$$

Пример 3. Ценность априорной независимости. Разбивая множество p признаков на два независимых подмножества (т.е. заменяя нулями выборочные корреляции признаков из разных подмножеств в предположении их независимости), с одной стороны, можно ухудшить решение задачи, так как предположение о независимости бывает обычно слабо обосновано, с другой стороны, улучшить устойчивость решения. Оценим это улучшение для трех оценок произвольной компоненты μ центра \mathbf{m} распределения $N_p(\mathbf{m}, \mathbf{I})$: ОМП $\hat{\mu} = \bar{x}$ в распределении, усеченном на уровне $|x - \mu| = 3$ по всем координатам, корневой оценки $\tilde{\mu}$ и ОМУ $\hat{\mu}^*$. Обозначим через W_1 вариационную производную дисперсии оценки $\hat{\mu}$ при оценке всех корреляций, через

Таблица 2

p	2	3	5	7	9	11
$\omega_p(\overset{\circ}{\mu})$	6,0	8,3	18	42	109	289
$\omega_p(\tilde{\mu})$	3,8	5,0	9,6	20	47	111
$\omega_p(\overset{*}{\mu})$	3,6	4,6	8,8	18	42	96

W_2 – вариационную производную при разбиении множества признаков на два независимых подмножества, комбинаторно осредненную по всевозможным разбиениям, а через $\omega_p(\tilde{\mu}) = W_1/W_2$ – их отношение. Тогда при $p = 2$ $\omega_2(\tilde{\mu}) = \tau(\tilde{\mu})$ (где $\tau(\overset{\circ}{\mu}) \approx 6$, $\tau(\tilde{\mu}) = 3\sqrt{\pi/2} \approx 3,76$, $\tau(\overset{*}{\mu}) = 2\sqrt{\pi} \approx 3,55$), а для нечетных $p \geq 3$

$$\omega_p = (2^{p-1} - 1) [(1 + \tau^{-1})^{p-1} - 1]^{-1}.$$

В результате получим табл. 2.

Вернемся к четырехмерным таксономическим задачам [4]. Популярность методов многомерной статистики привела к тому, что из публикаций исчезли первичные данные, приводятся лишь средние величины, так что можно анализировать задачу лишь умозрительно. Интерполируя в табл. 2, получим, что для $p = 4$ $\omega_4(\overset{\circ}{\mu}) \approx 12$, так что, отказываясь от разделения четырех признаков на два “независимых” подмножества (если оно не очень вредно для решения), увеличим вариационную производную во столько раз, во сколько она увеличилась бы при сокращении выборки в 12 раз, а для задачи [6] $n = 50$ и $50 : 12 \approx 4$ – слишком маленькая выборка. В этом, по-видимому, и кроется причина успеха чужеродных алгоритмов распознавания зрительных образов, игнорирующих связи признаков, в соревновании с многомерной статистикой, оценивающей эти связи аккуратно, но неустойчиво.

Для корневой оценки отношения $\omega_p(\tilde{\mu})$ в табл. 2 также растут, но медленнее, чем $\omega_p(\overset{\circ}{\mu})$. По-видимому, $p = 4$, $\omega_4(\tilde{\mu}) \approx 7$, это предельная (или уже запредельная) размерность для оценки $\tilde{\mu}$: при большей размерности надо делить признаки на “независимые” подмножества. Алгоритмы выделения “корреляционных плеяд” приводятся в [18], основаны они на максимально правдоподобной матрице корреляций, которую легко заменить компромиссной диагональной. Часто корреляция исчезает после удачной нормировки признаков.

Еще один способ сохранения устойчивости оценок – это сведение многомерной задачи к последовательности маломерных задач типа предложенного в [3] сведения многомерной дискриминации к последовательности двумерных задач.

Отношение $\omega_p(\overset{*}{\mu})$ для наиболее устойчивой оценки $\overset{*}{\mu}$ в табл. 2 показывает, что потеря устойчивости с ростом размерности является неизбежной. Статистический смысл этого вывода можно понять, рассмотрев “загрязняющее” наблюдение на периферии распределения. Так как плотность убывает как $(2\pi e)^{-p/2}$, вред от этого загрязнения растет как показательная функция p , различие рассмотренных оценок лишь в основаниях τ . Поэтому многомерная статистика не может иметь устойчивые оценки, они могут быть лишь у маломерной статистики.

Последний пример показывает, что многомерное нормальное распределение оказывается объектом непознаваемым: неустойчивость даже самой устойчивой ОМУ катастрофически растет с ростом размерности, потому что Лебегова мера носителя растет с ростом размерности. Это побуждает отказаться от этой общепринятой модели. Она не имеет каких-либо теоретических обоснований, так же как и

способов эмпирической проверки. Психологической основой модели является похожесть одномерных распределений на нормальные. Но одномерные распределения, похожие на нормальные, могут образовываться при проецировании многомерных распределений, далеких от многомерного нормального. Ниже предлагается модель многомерного распределения с ограниченным носителем, имеющая асимптотически нормальные проекции.

3. Сферическое распределение и его проекции

Пусть $v_p(r)$ – объем p -мерной сферы радиуса r , $v_p = v_p(1)$, а r_p – такой радиус, что $v_p(r_p) = 1$. Сферу Q_p с радиусом r_p (т.е. единичной массы) и центром в начале координат будем называть стандартной сферой.

Определения 1 и 2. Равномерное распределение наблюдений в стандартной p -мерной сфере Q_p будет называться стандартным сферическим распределением. Его проекция на прямую R^1 будет называться стандартным проекционным распределением.

Ниже при выписывании асимптотических функций при больших размерностях p знак \approx будет использоваться при переходе к главным частям таких функций, а знак \sim как символ асимптотической эквивалентности.

Теорема 2. Плотность стандартного проекционного распределения

$$f_p(z) = \begin{cases} c_p(1 - z^2/r_p^2)^{(p-1)/2} & \text{при } |z| \leq r_p, \\ 0 & \text{при } |z| \geq r_p, \end{cases}$$

где

$$c_p = [\Gamma(p/2 + 1)]^{1-1/p} / \Gamma((p-1)/2 + 1) \approx \sqrt{e},$$

$$r_p = \frac{1}{\sqrt{\pi}} [\Gamma(p/2 + 1)]^{1/p} \approx \sqrt{p/2\pi e} \sim \sqrt{p};$$

на конце распределения, при $z = -r_p$, ее производная составляет

$$\lim_{z \rightarrow -r_p} \frac{f_p(z)}{z + r_p} = \begin{cases} \infty, & p = 1, 2, \\ \sqrt[3]{6}\pi^{2/3}, & p = 3, \\ 0, & p \geq 4; \end{cases}$$

при $p \rightarrow \infty$ и малых значениях z/r_p распределение сходится к нормальному $N(0, 1/2\pi e)$.

Сходимость проекционного распределения к нормальному очень быстрая, это можно видеть на рис. 2, где приведены функции распределения для $p = 1, 2, 4$ и предельное нормальное распределение ($p = \infty$). Если учесть, что “тяжелые хвосты” и усеченность часто отмечаются при сравнении эмпирических распределений с нормальным, то проекционное распределение при небольших p может лучше описывать эмпирические распределения, похожие на нормальное, чем привычная модель нормального распределения.

Эмпирическое распределение можно рассматривать как линейную трансформацию стандартного проекционного распределения, когда вместо переменной z измеряется переменная x , такая, что $z = (x - \mu)/\omega$, и кроме размерности p надо оценить параметры локализации μ и масштаба ω . Чтобы понять, какие оценки могут здесь быть полезны, рассмотрим сначала случай, когда распределение известно с точностью до параметра локализации μ , и сравним три оценки для μ : ОМП $\hat{\mu}$, корневую оценку $\tilde{\mu}$ и полусумму $\bar{\mu} = (x_{(1)} + x_{(n)})/2$ крайних значений $x_{(1)}$ и $x_{(n)}$, назовем эту оценку экстремальной.

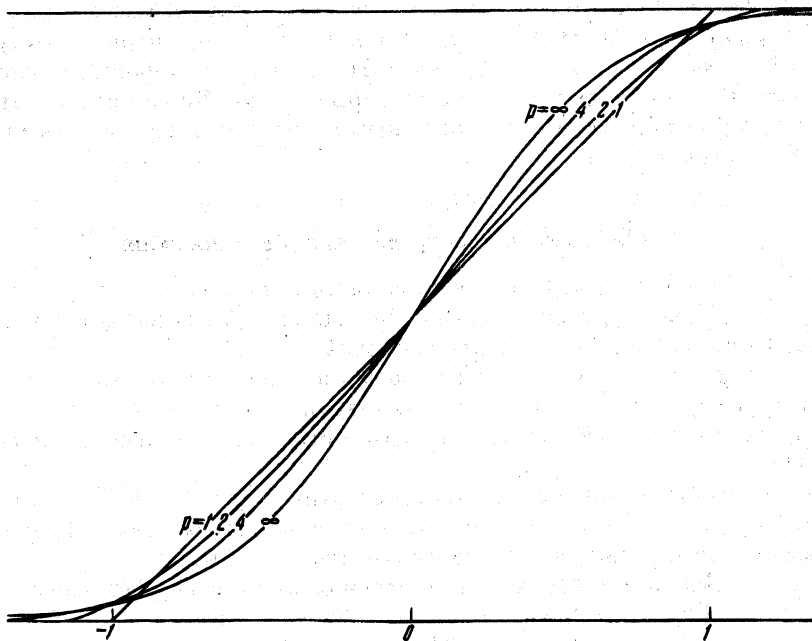


Рис. 2. Сравнение функций распределения стандартных проекционных распределений ($p = 1, 2, 4$) и стандартного нормального распределения ($p = \infty$)

Теорема 3. Для проекционного распределения, определенного с точностью до центра μ , $y = x - \mu$, $\omega = 1$,

1) ОМП $\hat{\mu}$ существует при $p \geq 4$ с оценочной функцией

$$\psi_p(y) = y(1 - y^2/r_p^2)^{-1},$$

$$\lim n E (\hat{\mu} - \mu)^2 = \begin{cases} \frac{2}{9\pi} r_4 & \text{при } p = 4, \\ \frac{r_p}{c_p} \frac{p-2}{(p-1)^2} c_{p-4} r_{p-4}^{p-4} \approx \frac{1}{2\pi e} & \text{при } p \geq 5, \end{cases}$$

но имеет нулевую устойчивость,

2) экстремальная оценка $\bar{\mu}$ существует при $p \geq 1$ и имеет

$$\lim n^{4/(p+1)} E (\bar{\mu} - \mu)^2 = \frac{1}{2} \left(\frac{r}{p+1} \right)^{2(p-1)/(p+1)} c_p^{-4/(p+1)} \approx (4\pi e p)^{-1}.$$

Прокомментируем результаты, выписанные в теореме. ОМП $\hat{\mu}$ и экстремальная оценка $\bar{\mu}$ имеют разный порядок убывания по n . Просчитывая дисперсии при разных n и p , можно убедиться, что для “реальных” сочетаний экстремальная оценка имеет обычно меньшую дисперсию, чем ОМП. Преимуществом экстремальной оценки является ее устойчивость к изменениям п.р. в интервале $(-r_p, r_p)$. Но загрязнение вне этого интервала может испортить оценку. “Грубое загрязнение” можно выбирать вручную или использовать формальную процедуру, предложенную в [13].

Масштабный параметр ω можно оценивать одновременно с центром μ , сравнивая на дисплее линейно деформированную эмпирическую функцию распределения с функциями распределения стандартных проекционных распределений для различных размерностей p .

Перейдем к совместному распределению нескольких переменных. Пусть они уже центрированы и нормированы, так что их можно считать стандартными. Тогда наблюдения в различных переменных можно рассматривать как проекции одной выборки из сферического распределения на прямые, разноориентированные в R^p , и задача сводится к оценке косинусов углов между этими прямыми, т.е. парных корреляций ρ переменных. По аналогии с теоремой 3 экстремальные оценки следует предпочесть оценкам минимума контраста. При коррелированности двух признаков на плоскости, соответствующей этим признакам, наблюдения образуют облако точек, обозначим через d_1 размах проекции этого облака на биссектрису координатного угла, а через d_2 размах проекции на нормаль к биссектрисе, и пусть $\hat{\ell} = d_1^2/d_2^2$. Собственные значения λ_1 и λ_2 матрицы корреляций удовлетворяют уравнению

$$\left| \begin{pmatrix} \ell & \rho \\ \rho & \ell \end{pmatrix} - \lambda \mathbf{I} \right| = (1 - \lambda^2) - \rho^2 = 0,$$

откуда $1 - \lambda = \pm |\rho|$, так что $\lambda_1 = 1 + |\rho|$, $\lambda_2 = 1 - |\rho|$, их отношение $1 = \lambda_1/\lambda_2$ и $|\rho| = (\ell - 1)/(\ell + 1)$. Заменив ℓ на $\hat{\ell}$, мы получили экстремальную оценку для $|\rho|$. Оценив все парные корреляции, получим оценку матрицы корреляций R , с которой можем делать процедуры, рекомендуемые классической многомерной статистикой, полученные при рассмотрении первых двух моментов распределения и не использующих нормальности на периферии распределения. Если в сферическое распределение поверить так же, как мы верили в многомерную нормальность, то становится предельно простым отбор признаков: для характеристики p -мерного сферического распределения достаточно взять p признаков, а остальные отбросить.

4. Выводы

Неудачи многомерной статистики в соревновании с алгоритмами распознавания образов определяются неустойчивостью используемых оценок максимума правдоподобия. Вариационная производная нормированной дисперсии оценки центра нормального распределения, усеченного на уровне 3σ , при добавлении каждого нового признака возрастает почти в 6 раз. Переход к более устойчивым оценкам несколько улучшает положение: для наиболее устойчивой оценки добавление нового признака увеличивает вариацию почти в 3 раза, так что существенно многомерное нормальное распределение оказывается непознаваемым. Для получения устойчивых оценок можно множество признаков разделить на "независимые" подмножества небольшой размерности (вряд ли больше двух-трех признаков в каждом) и использовать корневые или компромиссные оценки. Второй путь получения устойчивых оценок — это замена модели многомерного нормального распределения моделью равномерного распределения в многомерной сфере единичного объема, в которой наблюдаемые признаки можно рассматривать как линейно деформированные проекции на разноориентированные прямые.

5. Доказательства

Лемма. Оценка \hat{C} эквивариантна [15], поэтому для доказательства достаточно рассмотреть случай, когда $\mathbf{m} = \mathbf{0}$, $\mathbf{C} = \mathbf{I}$. Для состоятельной оценки \hat{C} должно выполняться равенство $E\psi(\mathbf{x}, \mathbf{C}) = \mathbf{0}$ для всех функций ψ , т.е.

$$(7) \quad E \mathbf{x} \mathbf{x}^T \eta(\mathbf{x}^T \mathbf{x}) = \beta E \eta(\mathbf{x}^T \mathbf{x}) \mathbf{I}.$$

Пусть $p \geq 2$. Обозначая через \int_{R^p} интеграл по пространству R^p , а через S_p поверхность p -мерной сферы единичного радиуса, найдем правую часть (7):

$$\begin{aligned} E\eta(\mathbf{x}^T \mathbf{x}) &= (2\pi)^{-p/2} \int_{R^p} \eta(\mathbf{x}^T \mathbf{x}) \exp(-\mathbf{x}^T \mathbf{x}/2) d\mathbf{x} = \\ &= (2\pi)^{-p/2} S_p \int_0^\infty r^{p-1} \eta(r^2) \exp(-r^2/2) dr = \\ &= \frac{1}{2} \pi^{-p/2} S_p \int_0^\infty s^{p/2-1} \eta(2s) \exp(-s) ds; \end{aligned}$$

здесь произведены замены переменных: $r^2 = \mathbf{x}^T \mathbf{x}$, $s = r^2/2$. Используя те же замены переменных, найдем функцию S_p из равенства единице интеграла по п.р. $N_p(0, \mathbf{I})$:

$$\begin{aligned} 1 &= (2\pi)^{-p/2} \int_{R^p} \exp(-\mathbf{x}^T \mathbf{x}/2) d\mathbf{x} = (2\pi)^{-p/2} S_p \int_0^\infty r^{p-1} \exp(-r^2/2) dr = \\ &= \frac{1}{2} \pi^{-p/2} S_p \int_0^\infty s^{p/2-1} \exp(-s) ds = \frac{1}{2} \pi^{-p/2} S_p \Gamma(p/2), \end{aligned}$$

$$(8) \quad S_p = 2\pi^{p/2} / \Gamma(p/2), \quad p \geq 2,$$

так что

$$(9) \quad E\eta(\mathbf{x}^T \mathbf{x}) = \int_0^\infty s^{p/2-1} \eta(2s) \exp(-s) ds / \Gamma(p/2).$$

Чтобы найти левую часть (7), разделим случайное событие \mathbf{x} на следующие три независимых случайных действия. 1) В пространстве R^p случайно выберем ортогональную систему координат, матрицу направляющих косинусов обозначим через \mathbf{H} , а математическое ожидание по этому случайному событию — через E^H . 2) В p -мерном векторе случайно выберем одну компоненту и положим ее равной единице, а остальные компоненты приравняем нулю; полученный случайный вектор обозначим через \mathbf{u} , а математическое ожидание по этому случайному событию — через E^u . 3) Введем случайную величину $q = \mathbf{x}^T \mathbf{x}$, математическое ожидание по этой случайной величине обозначим через E^q . Тогда проекция вектора $\mathbf{H}\sqrt{q}\mathbf{u}$ равномерно распределена на поверхности сферы с центром в начале координат, длина вектора равна \sqrt{q} , поэтому $\mathbf{H}\sqrt{q}\mathbf{u} \sim N_p(0, \mathbf{I})$. Теперь можем найти левую часть (7):

$$\begin{aligned} E\mathbf{x}\mathbf{x}^T(\mathbf{x}^T \mathbf{x}) &= E\mathbf{H}\sqrt{q}\mathbf{u}(\mathbf{H}\sqrt{q}\mathbf{u})^T \eta(q) = E^q E^H E^u \mathbf{H} q \mathbf{u} \mathbf{u}^T \mathbf{H}^T \eta(q) = \\ &= \frac{1}{p} E^q q \eta(q) E^H \mathbf{H} \mathbf{H}^T = \frac{1}{p} E^q q \eta(q) \mathbf{I} = \frac{1}{p} E q \eta(q) \mathbf{I}; \end{aligned}$$

здесь при переходе к последнему равенству у оператора E убран уже не нужный индекс q . Найдем множитель при матрице \mathbf{I} в полученном выражении, используя тот же прием и те же замены переменных, что и при получении равенства (9):

$$\begin{aligned}\frac{1}{p} E q \eta(q) &= \frac{1}{p} (2\pi)^{-p/2} S_p \int_0^\infty r^{p+1} e^{-r^2/2} \eta(r^2) dr = \\ &= \frac{1}{p} \pi^{-p/2} S_p \int_0^\infty s^{p/2} e^{-s} \eta(2s) ds = \frac{2}{p} \int_0^\infty s^{p/2} e^{-s} \eta(2s) ds / \Gamma(p/2).\end{aligned}$$

Подставив полученное выражение и (9) в (7), получим функцию $\beta(p)$, выписанную в лемме.

Пусть теперь $\beta(p) = \beta$ не зависит от p . Весовая функция $\eta(q)$ определена с точностью до ненулевого множителя, выберем его таким, что для п.р. $\varphi(s)$ распределения $N_1(0, 1) \int_{-\infty}^\infty \eta(s^2) \varphi(s) ds = 1$. Тогда функция $g(s) = \eta(s^2) \varphi(s)$ может рассматриваться как некоторая одномерная п.р. Она симметрична, поэтому имеет нулевые нечетные моменты. Ее p -й четный момент обозначим через ζ_p . Полученное выражение для $\beta(p)$ дает следующие отношения четных моментов:

$$\zeta_2/\zeta_0 = \beta, \quad \zeta_4/\zeta_2 = 3\beta, \dots, \zeta_{2k}/\zeta_{2(k-1)} = (2k-1)\beta, \dots$$

Очевидно, $\zeta_0 = 1$. Последовательно находим:

$$\zeta_2 = \beta, \quad \zeta_4 = 1 \cdot 3\beta^2, \dots, \zeta_{2k} = 1 \cdot 3 \cdot 5 \dots (2k-1)\beta^k, \dots$$

Но это четные моменты распределения $N(0, \beta)$. Следовательно,

$$g(s) = \eta(s^2) \varphi(s) = (2\pi\beta)^{-1/2} \exp(-s^2/2\beta),$$

и с точностью до несущественного множителя $(2\pi\beta)^{-1/2}$ получим экспоненциальную весовую функцию с $\lambda = \beta^{-1} - 1$, т.е. $\beta = (1 + \lambda)^{-1}$.

Обратное утверждение о независимости β от p при экспоненциальной весовой функции доказано в [10].

Теорема 1. Система уравнений для условно-оптимальных оценок $\tilde{\mathbf{m}}$ и $\tilde{\mathbf{C}}$ следует из теоремы 2 из [16], и остается лишь найти эффективность и устойчивость оценки $\tilde{\mu}$. Сделаем это, используя тот же прием и те же замены переменных, что и при доказательстве леммы.

Оценка \mathbf{m} эквивариантна, поэтому можно искать величины $\text{eff } \tilde{\mu}$ и $\text{stb } \tilde{\mu}$ для распределения $N_p(\mathbf{m}, \mathbf{I})$, плотность которого обозначим через $\varphi_p(\mathbf{y})$, $\mathbf{y} = \mathbf{x} - \mathbf{m}$.

При $p \geq 2$ для произвольной компоненты μ вектора \mathbf{m} соответствующие компоненты векторов \mathbf{x} и \mathbf{y} обозначим через a и b . Для компромиссной оценки $\tilde{\mu}$ параметра μ используется функция

$$\psi(a - \mu) = (a - \mu) [1 + \gamma/\varphi_p(\mathbf{y})]^{-1}.$$

Из леммы в [16] следует, что

$$(10) \quad \text{eff } \tilde{\mu} = E \psi^2(a - \mu) \left[E \frac{d}{d\mu} \psi(a - \mu) \right]^{-2}.$$

Упростим выражение в квадратных скобках. Интегрируя по частям, так же как в примере 3 [16], получим, что

$$\begin{aligned} E \frac{d}{d\mu} \psi(a - \mu) &= \int_{R^p} \left[\frac{d}{d\mu} \psi(a - \mu) \right] \varphi_p(y) dy = \int_{R^p} \psi(a - \mu) \frac{d}{db} \varphi_p(y) dy = \\ &= - \int_{R^p} \psi(a - \mu)(a - \mu) \varphi_p(y) dy = - \int_{R^p} b^2 [1 + \gamma / \varphi_p(y)]^{-1} \varphi_p(y) dy. \end{aligned}$$

Просуммировав производные по всем компонентам вектора \mathbf{m} и поделив результат на p , получим, что

$$\begin{aligned} (11) \quad E \frac{d}{d\mu} \psi(a - \mu) &= - \frac{1}{p} \int_{R^p} \mathbf{y}^T \mathbf{y} [1 + \gamma / \varphi_p(y)]^{-1} \varphi_p(y) dy = \\ &= - \frac{1}{p} (2\pi)^{-p/2} S_p \int_0^\infty r^{p+1} [1 + \gamma (2\pi)^{p/2} e^{r^2/2}]^{-1} e^{-r^2/2} dr = \\ &= - \int_0^\infty s^{p/2} [1 + \gamma (2\pi)^{p/2} e^s]^{-1} e^{-s} ds / \Gamma(p/2 + 1). \end{aligned}$$

Аналогично находим

$$(12) \quad E \psi^2(a - \mu) = \int_0^\infty s^{p/2} [1 + \gamma (2\pi)^{p/2} e^s]^{-2} e^{-s} ds / \Gamma(p/2 + 1).$$

Подставив (11) и (12) в (10), получим значение $\text{eff } \tilde{\mu}$, выписанное в теореме.

Для $p = 1$, интегрируя, находим, что

$$\begin{aligned} E \frac{d}{d\mu} \psi(x - \mu) &= \frac{2}{\sqrt{\pi}} \int_0^\infty \sqrt{s} e^{-s} (1 + \gamma \sqrt{2\pi} e^s)^{-1} ds, \\ E \psi^2(x - \mu) &= \frac{2}{\sqrt{\pi}} \int_0^\infty \sqrt{s} e^{-s} (1 + \gamma \sqrt{2\pi} e^s)^{-2} ds, \\ \text{eff } \tilde{\mu} &= \frac{2}{\sqrt{\pi}} \left[\int_0^\infty \sqrt{s} e^{-s} (1 + \gamma \sqrt{2\pi} e^s)^{-1} ds \right]^2 / \int_0^\infty \sqrt{s} e^{-s} (1 + \gamma \sqrt{2\pi} e^s)^{-2} ds, \end{aligned}$$

так что формула $\text{eff } \tilde{\mu}$ в теореме верна для всех p .

Для нахождения устойчивости нужно значение минимальной вариационной производной W^* . Для ОМУ

$$\begin{aligned}
 \psi^*(a - \mu) &= \frac{d}{d\mu} \exp(-\mathbf{y}^T \mathbf{y} / 2) = (a - \mu) \exp(-\mathbf{y}^T \mathbf{y} / 2); \\
 E \frac{d}{d\mu} \psi^*(a - \mu) &= (2\pi)^{-p/2} \int_{R^p} b^2 \exp(-\mathbf{y}^T \mathbf{y}) d\mathbf{y} = 2^{-p/2-1}, \\
 \int_{R^p} \psi^{*2}(a - \mu) d\mathbf{x} &= \int_{R^p} b^2 \exp(-\mathbf{y}^T \mathbf{y}) d\mathbf{y} = \frac{1}{2} \pi^{p/2}, \\
 W^* &= \int_{R^p} \psi^{*2}(b) d\mathbf{y} \left[E \frac{d}{d\mu} \psi^*(b) \right]^{-2} = 2^{p+1} \pi^{p/2}.
 \end{aligned}
 \tag{13}$$

Для $p \geq 2$

$$\begin{aligned}
 \int_{R^p} \psi^2(b) d\mathbf{y} &= \frac{1}{p} \int_{R^p} \mathbf{y}^T \mathbf{y} \left[1 + \gamma (2\pi)^{p/2} \exp(\mathbf{y}^T \mathbf{y} / 2) \right]^{-2} d\mathbf{y} = \\
 &= (2\pi)^{p/2} \int_0^\infty s^{p/2} \left[1 + \gamma (2\pi)^{p/2} e^s \right]^{-2} ds / \Gamma(p/2 + 1),
 \end{aligned}$$

что дает формулу устойчивости, выписанную в теореме. Для $p = 1$

$$\int_{-\infty}^\infty \psi^2(b) db = 2\sqrt{2} \int_0^\infty \sqrt{s} \left[1 + \gamma \sqrt{2\pi} e^s \right]^{-2} ds,$$

так что формула устойчивости верна и для $p = 1$.

Для получения компромиссных оценок надо найти константы γ_1 и γ_2 в системе уравнений для условно-оптимальных оценок. Сделаем это, взяв за координаты собственные векторы матрицы \mathbf{C} . При повороте системы координат сохраняются значения $f(x_i)$ в знаменателях и равенство нулю суммы линейных функций от наблюдений в первом уравнении системы и суммы квадратичных форм во втором уравнении. Ставшую диагональной матрицу ковариаций обозначим через \mathbf{D} . Обозначение a сохраним для произвольной компоненты вектора \mathbf{x} , так же как и $\mu = E a$ и $b = a - \mu$, а дисперсию этой компоненты обозначим через $\sigma^2 = E(a - \mu)^2$.

Найдем компромиссную оценку $\overset{\circ}{\mu}$ для \mathbf{x} , распределенного $N_p(\mathbf{m}, \mathbf{D})$, $\mathbf{y} = \mathbf{x} - \mathbf{m}$, считая матрицу \mathbf{D} известной. Для оценки $\overset{\circ}{\mu}$

$$\begin{aligned}
 \overset{\circ}{V} &= \sigma^2, \\
 \psi^*(a - \mu) &= b \exp(-\mathbf{y}^T \mathbf{D}^{-1} \mathbf{y} / 2), \\
 \frac{d}{d\mu} \psi^*(a - \mu) &= -b^2 \sigma^{-2} \exp(-\mathbf{y}^T \mathbf{D}^{-1} \mathbf{y} / 2), \\
 E \frac{d}{d\mu} \psi^*(a - \mu) &= -(2\pi)^{-p/2} |\mathbf{D}|^{-1/2} \sigma^{-2} \int_{R^p} b^2 \exp(-\mathbf{y}^T \mathbf{D}^{-1} \mathbf{y}) d\mathbf{y} = -2^{-p/2-1}, \\
 \int_{R^p} \psi^{*2}(a - \mu) d\mathbf{y} &= \int_{R^p} b^2 \exp(-\mathbf{y}^T \mathbf{D}^{-1} \mathbf{y}) d\mathbf{y} = \frac{1}{2} \pi^{p/2} |\mathbf{D}|^{1/2} \sigma^2, \\
 W^* &= 2^{p+1} \pi^{p/2} |\mathbf{D}|^{1/2} \sigma^2,
 \end{aligned}$$

$$\gamma_1 = \dot{V} / \dot{W} = 2^{-p-1} \pi^{-p/2} |\mathbf{D}|^{-1/2},$$

$$\gamma_1 / f(\mathbf{x}) = 2^{-p/2-1} \exp(-\mathbf{y}^T \mathbf{D}^{-1} \mathbf{y} / 2),$$

и получаем первое уравнение системы для компромиссных оценок.

Для оценки $\dot{\sigma}$

$$\begin{aligned} \dot{V} &= \sigma^2 / 2, \\ \dot{\psi}(b/\sigma) &= (a^2 \sigma^{-2} - 1/2) \exp(-\mathbf{y}^T \mathbf{D}^{-1} \mathbf{y} / 2), \\ \frac{d}{d\sigma} \dot{\psi}(b/\sigma) &= \sigma^{-1} (a^4 \sigma^{-4} - \frac{5}{2} a^2 \sigma^{-2}) \exp(-\mathbf{y}^T \mathbf{D}^{-1} \mathbf{y} / 2), \\ E \frac{d}{d\sigma} \dot{\psi}(b/\sigma) &= (2\pi)^{-p/2} |\mathbf{D}|^{-1/2} \sigma^{-1} \int_{R^p} \left(a^4 \sigma^{-4} - \frac{5}{2} a^2 \sigma^{-2} \right) \exp(-\mathbf{y}^T \mathbf{D}^{-1} \mathbf{y}) d\mathbf{y} = \\ &= -2^{-p/2-1} \sigma^{-1}, \\ \int_{R^p} \dot{\psi}^2(b/\sigma) d\mathbf{y} &= \int_{R^p} \left(a^4 \sigma^{-4} - a^2 \sigma^{-2} + \frac{1}{4} \right) \exp(-\mathbf{y}^T \mathbf{D}^{-1} \mathbf{y}) d\mathbf{y} = \frac{1}{2} \pi^{p/2} |\mathbf{D}|^{1/2}, \\ \dot{W} &= 2^{p+1} \pi^{p/2} |\mathbf{D}|^{1/2} \sigma^2, \\ \gamma_2 = \dot{V} / \dot{W} &= 2^{p+2} \pi^{p/2} |\mathbf{D}|^{1/2}, \\ \gamma_2 / f(\mathbf{x}) &= 2^{-p/2-2} \exp(-\mathbf{y}^T \mathbf{D}^{-1} \mathbf{y} / 2), \end{aligned}$$

и получаем второе уравнение системы для компромиссных оценок.

Корневые оценки $\tilde{\mathbf{m}}$ и $\tilde{\mathbf{C}}$ выписываются элементарно. Значение $\beta = 2/3$ во втором уравнении системы следует из леммы при $\lambda = 1/2$.

Пример 1. Формула (6) выводится в [20], 42.2. Как и в [17], в [20] избавляются от смещения $O(n^{-1})$ в определителе, но другим, более сложным способом.

Найдем смещение при упрощенном условии, что центр распределения известен. Пусть $\mathbf{I} + \Delta$, $\Delta = \{\delta_{ij}\}$ – матрица выборочных ковариаций для распределения $N_p(0, \mathbf{I})$. Тогда $\dot{\mathbf{C}} = \sqrt{\mathbf{C}}(\mathbf{I} + \Delta)\sqrt{\mathbf{C}}$ и

$$(14) \quad E |\dot{\mathbf{C}}| = E |\sqrt{\mathbf{C}}| |\mathbf{I} + \Delta| |\sqrt{\mathbf{C}}| = |\mathbf{C}| E |\mathbf{I} + \Delta| = |\mathbf{C}| [1 - p(p-1)/2n + O(n^{-2})],$$

так как в определителе $|\mathbf{I} + \Delta|$ величинами $O(n^{-1})$ являются $p(p-1)/2$ слагаемых вида $-\delta_{ij}\delta_{ji}$, математическое ожидание каждого из них равно $-1/n$. Полученное смещение отличается от (6) заменой $p+1$ на $p-1$ из-за предположения об известном центре распределения.

При том же предположении рассмотрим диагональную корневую оценку, сохранив обозначение $\mathbf{I} + \Delta$ для корневой оценки в распределении $N_p(0, \mathbf{I})$ и считая, что $\nu = O(n^{-1})$:

$$\begin{aligned} E \left| (1-\nu)\tilde{\mathbf{C}} + \nu\tilde{\mathbf{C}}_{dg} \right| &= E \left| \sqrt{\mathbf{C}} \right| \left| (1-\nu)(\mathbf{I} + \Delta) + \nu\mathbf{C}^{-1/2}\tilde{\mathbf{C}}_{dg}\mathbf{C}^{-1/2} \right| \left| \sqrt{\mathbf{C}} \right| = \\ &= |\mathbf{C}| E \left| (1-\nu)(\mathbf{I} + \Delta) + \mathbf{C}^{-1/2}\tilde{\mathbf{C}}_{dg}\mathbf{C}^{-1/2} \right|. \end{aligned}$$

Здесь, кроме $p(p-1)/2$ слагаемых вида $-\delta_{ij}\delta_{ji}$, величинами $O(n^{-1})$ будут являться p слагаемых, равных $-\nu$, и p слагаемых, составляющих $\nu \operatorname{tr} (\mathbf{C}^{-1/2}\tilde{\mathbf{C}}\mathbf{C}^{-1/2})$. В [11, 12] показано, что $E \delta_{ij}^2 = (3/2\sqrt{2})^{p+4}/n$. В величине $\nu \operatorname{tr} (\mathbf{C}^{-1/2}\tilde{\mathbf{C}}_{dg}\mathbf{C}^{-1/2}) =$

$= O(n^{-1})$ замена \tilde{C} на C приведет к изменениям в членах более высокого порядка малости, чем $O(n^{-1})$; равенство $\text{tr} (C^{-1/2} C_{\text{dg}} C^{-1/2}) = \text{tr} (C_{\text{dg}}^{-1/2} C C_{\text{dg}}^{-1/2}) = \text{tr} R^{-1}$ проверяется непосредственно. Получаем, что

$$E \left| (1-\nu)\tilde{C} + \nu C_{\text{dg}} \right| = |C| \left[1 - \frac{p(p-1)}{2n} (3/2\sqrt{2})^{p+4} + \nu(\text{tr} R^{-1} - p) + O(n^{-2}) \right],$$

и для обусловленности оценки должно выполняться равенство

$$\nu = (3/2\sqrt{2})^{p+4} p(p-1) [2n(\text{tr} R^{-1} - p)]^{-1}.$$

Делая замену $p-1$ на $p+1$, как при переходе от (14) к (6), получаем коэффициент ν для диагональной компромиссной оценки, выписанный в примере.

Пример 2. Доказательство в [19].

Пример 3. Заменяем $E\psi$ в усеченном распределении на $E\psi = 1$ в распределении $N(0, 1)$. Тогда вариационная производная дисперсии среднего \bar{x}_1 составит

$$(15) \quad W(p) \approx \int_{-3}^3 y_1^2 dy_1 \int_{-3}^3 dy_2 \dots \int_{-3}^3 dy_p = 2^p \int_0^3 y_1^2 dy_1 \int_0^3 dy_2 \dots \int_0^3 dy_p = 3\tau^p(\dot{\mu}),$$

где $\tau(\dot{\mu}) = 6$. Для $p = 2$

$$\omega_2(\dot{\mu}) = W_1/W_2 = W(2)/W(1) = \tau(\dot{\mu}) = \tau.$$

Для нечетного $p \geq 3$ существует C_p^q способов деления множества признаков на подмножества с $q \leq (p-1)/2$ и $p-q \geq (p-1)/2$ признаками, при каждом из которых сумма по признакам вариационных производных составит $3[q\tau^q + (p-q)\tau^{p-q}]$. Сумма вариационных производных по всевозможным разделениям составит сумму

$$\begin{aligned} 3 \sum_{q=1}^{(p-1)/2} [q\tau^q + (p-q)\tau^{p-q}] C_p^q &= 3 \sum_{q=1}^{p-1} q\tau^q C_p^q = 3 [p\tau(1+\tau)^{p-1} - p\tau^p] = \\ &= 3p\tau [(1+\tau)^{p-1} - \tau^{p-1}] \end{aligned}$$

([21], 4.2.3.18), при этом количество просуммированных вариационных производных составит сумму

$$\sum_{q=1}^{p-1} q C_p^q = 2^{p-1} p - p = p(2^{p-1} - 1)$$

([21], 4.2.2.7), так что осредненная вариационная производная равна

$$W_2 \approx 3\tau [(1+\tau)^{p-1} - \tau^{p-1}] (2^{p-1} - 1)^{-1}.$$

Разделив $W_1 = W(p)$ (15) на W_2 и поделив числитель и знаменатель на τ^{p-1} , получим отношение $\omega_p(\dot{\mu})$, выписанное в примере.

Рассмотрение корневых оценок проводится вполне аналогично, различие лишь в величине $W(p)$, которая составляет

$$\begin{aligned} W(p) &= \int_{R^p} b^2 \exp(-\mathbf{y}^T \mathbf{y}/2) d\mathbf{y} \left[(2\pi)^{-p/2} \int_{R^p} b^2 \exp(-3\mathbf{y}^T \mathbf{y}/4) d\mathbf{y} \right]^{-2} = \\ &= (2\pi)^{p/2} \left[(2/3)^{p/2+1} \right]^{-2} = (9/4)(3\sqrt{\pi/2})^p = (9/4)\tau^p(\tilde{\mu}). \end{aligned}$$

Для ОМУ μ^* значение $W(p) = \dot{W} = 2(2\sqrt{\pi})^p = 2\tau^p(\mu^*)$ выписано в (13).

Теорема 2. Используя формулу (8), найдем объем сферы единичного радиуса:

$$v_p = v_p(1) = S_p(1) \int_0^1 r^{p-1} dr = \frac{1}{p} S_p(1) = \pi^{p/2} / \Gamma(p/2 + 1).$$

Для сферы с единичной массой $v_p(r_p) = 1$,

$$r_p = v_p^{-1/p} = [\Gamma(p/2 + 1)]^{1/p} / \sqrt{\pi}.$$

Спроектируем эту сферу на произвольную прямую с переменной $z \in Z = R^1$. Величина $f_p(z)$ равна объему $(p-1)$ -мерной сферы с радиусом $\sqrt{r_p^2 - z^2}$, т.е.

$$f_p(z) = (r_p^2 - z^2)^{(p-1)/2} v_{p-1} = r_p^{p-1} (1 - z^2/r_p^2)^{(p-1)/2} v_{p-1} = c_p (1 - z^2/r_p^2)^{(p-1)/2},$$

$$c_p = r_p^{p-1} v_{p-1} = [\Gamma(p/2 + 1)]^{1-1/p} / \Gamma((p-1)/2 + 1).$$

В теореме выписана эта плотность и после знака \approx указан главный член при больших p , полученный по формуле Стирлинга.

Найдем производную этой плотности в точке $z = -r_p$:

$$\lim_{z \rightarrow -r_p} \frac{f_p(z)}{r_p + z} = \lim \frac{(r_p - z) f_p(z)}{r_p^2 - z^2} = \lim \frac{(r_p^2 - z^2)^{(p-1)/2} v_{p-1}}{r_p^2 - z^2} (r_p - z) =$$

$$= 2r_p v_{p-1} (r_p^2 - z^2)^{(p-3)/2},$$

подставив в полученную формулу значения r_p и v_{p-1} , выписанные выше, получим значения производной, указанные в теореме.

Отыскивая предельное проекционное распределение при $p \rightarrow \infty$, получим, что $(1 - z^2/r_p^2)^{(p-1)/2}$ сходится к экспоненте и

$$f_p(z) \approx \sqrt{e} \exp\left(-\frac{1}{2} x^2 (p-1)/r_p^2\right) \approx \sqrt{e} \exp\left(-\frac{1}{2} x^2 / 2\pi e\right).$$

Теорема 3. Логарифмируя и дифференцируя п.р. $f_p(y)$, получим оценочную функцию $\dot{\psi}_p$ для ОМП $\dot{\mu}$, выписанную в теореме. При отыскании ее дисперсии используем для биномиального дифференциала равенство

$$\int_0^1 t^2 (1-t^2)^k dt = \frac{1}{2k+3} \int_0^1 (1-t^2)^k dt, \quad k \geq 0:$$

$$E \left[\frac{d}{d\mu} \ln f_p(y) \right]^2 = E \left[-\frac{p-1}{r_p^2} \frac{y}{1-y^2/r_p^2} \right]^2 =$$

$$= \frac{(p-1)^2}{r_p^4} c_p \int_{-r_p}^{r_p} y^2 (1-y^2/r_p^2)^{(p-5)/2} dy =$$

$$= \frac{2}{r_p} (p-1)^2 c_p \int_0^1 t^2 (1-t^2)^{(p-5)/2} dt = \frac{2}{r_p} \frac{(p-1)^2}{p-2} c_p \int_0^1 (1-t^2)^{(p-5)/2} dt.$$

Последний интеграл существует лишь при $p \geq 4$. При $p = 4$ он равен $\pi/2$. При $p \geq 5$

$$\begin{aligned} E \left[\frac{d}{d\mu} \ln f_p(y) \right]^2 &= 2 \frac{c_p}{r_p} \frac{(p-1)^2}{p-2} r_{p-4}^{-p+4} \int_0^1 (r_{p-4}^2 - r_{p-4}^2 t^2)^{(p-5)/2} dt = \\ &= \frac{c_p}{r_p} \frac{(p-1)^2}{p-2} r_{p-4}^{-p+4} \int_{-r_{p-2}}^{r_{p-2}} (r_{p-4}^2 - s^2)^{(p-5)/2} ds = \frac{c_p}{r_p} \frac{(p-1)^2}{p-2} \frac{1}{c_{p-4}} r_{p-4}^{-p+4}, \end{aligned}$$

потому что интеграл в предпоследнем выражении равен v_{p-4}/c_{p-4} . Обратная величина составит

$$\lim n E \left(\overset{\circ}{\mu} - \mu \right)^2 = \frac{r_p}{c_p} \frac{p-2}{(p-1)^2} c_{p-4} r_{p-4}^{p-4} \approx \frac{1}{2\pi e},$$

если воспользоваться формулой Стирлинга. Вариационная производная пропорциональна интегралу

$$\int_{-r}^r \psi^2(y) dy = \int_{-r}^r y^2 (1 - y^2/r^2) dy,$$

который расходится при всех p , отсюда следует, что устойчивость $\overset{\circ}{\mu} = 0$.

Для получения дисперсии экстремальной оценки найдем асимптотическое ($n \rightarrow \infty$) распределение для наименьшего элемента выборки $y_{(1)}$. В окрестности $y = -r_p$ обозначим $u = y + r_p$, будем использовать знак \simeq при выделении главных частей выражений при $n \rightarrow \infty$ и сокращенные обозначения $c = c_p$, $r = r_p$. Тогда функция распределения величины u составит

$$\begin{aligned} F(u) &= \int_0^u f_p(t-r) dt = c_p \int_0^u [1 - (t/r - 1)^2]^{(p-1)/2} dt \simeq \\ &\simeq c \left(\frac{2}{r} \right)^{(p-1)/2} \int_0^u t^{(p-1)/2} dt = \frac{2c}{p+1} \left(\frac{2}{r} \right)^{(p-1)/2} u^{(p+1)/2}. \end{aligned}$$

Известно (см., например, [8, 12]), что $E F(y_{(1)}) = 1/n$. Пусть u_1 есть решение уравнения $F(u_1) = 1/n$. Тогда

$$\begin{aligned} u_1 &= \left[\frac{p+1}{2cn} \left(\frac{r}{2} \right)^{(p-1)/2} \right]^{2/(p+1)} = \left(\frac{r}{2} \right)^{\frac{p-1}{p+1}} \left(\frac{p+1}{2cn} \right)^{\frac{2}{p+1}} = \frac{1}{2} r^{\frac{p-1}{p+1}} \left(\frac{p+1}{cn} \right)^{\frac{2}{p+1}}, \\ f_p(u_1 - r) &\simeq c \left[\frac{2}{r} \left(\frac{r}{2} \right)^{\frac{p-1}{p+1}} \left(\frac{p+1}{2cn} \right)^{\frac{2}{p+1}} \right]^{\frac{p-1}{2}} = c \left(\frac{p+1}{crn} \right)^{\frac{p-1}{p+1}}. \end{aligned}$$

Величина $F(y_{(1)})$ имеет стандартное отклонение $1/n + O(1/n)$, а величина $y_{(1)}$, в $f_p(u_1 - r)$ раз меньшее, т.е.

$$\sqrt{E(y_{(1)} - E y_{(1)})^2} = \frac{1}{cn} \left(\frac{crn}{p+1} \right)^{\frac{p-1}{p+1}} = \left(\frac{r}{p+1} \right)^{\frac{p-1}{p+1}} \frac{1}{(cn)^{2/(p+1)}}.$$

Возведя эту величину в квадрат и поделив пополам (так как $\bar{\mu}$ есть полусумма асимптотически независимых экстремальных значений), мы получим дисперсию оценки $\bar{\mu}$, выписанную в теореме.

СПИСОК ЛИТЕРАТУРЫ

1. Бонгард М. М. Проблемы узнавания. М.: Наука, 1967.
2. Шурыгин А. М. Сравнение эффективности различных алгоритмов классификации при решении реальных задач // Алгоритмическое и программное обеспечение прикладного статистического анализа. М.: Наука, 1980. С. 354–360.
3. Шурыгин А. М. Поиск наилучшего деления двух выборок // Статистические методы классификации. М.: Изд. МГУ, 1969. С. 53–57.
4. Anderson T. W. An introduction to multivariate statistical analysis. N.Y.: John Wiley and Sons, 1958. (Перевод: Андерсон Т. Введение в многомерный статистический анализ. М.: Физматгиз, 1963.)
5. Fisher R. A. The use of multiple measurements in taxonomic problems // Ann. Eugen. 1936. № 7. P. 179–188.
6. Rao C. R. A statistical criterion to determine the group to which an individual belongs // Nature. 1947. V. 160. P. 835–836.
7. Деев А. Д. Представление статистик дискриминантного анализа и асимптотическое разложение при размерности пространства, сравнимой с размером выборок // Докл. АН СССР. 1970. Т. 195. С. 759–762.
8. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983.
9. Huber P. J. Robust Statistics. N.Y.: John Wiley and Sons, 1981. (Перевод: Хьюбер П. Робастность в статистике. М.: Мир, 1984.)
10. Meshalkin L. D. Some mathematical methods for the study of noncommunicable diseases // Proc. 6th Intern. Meeting on Uses of Epidemiol. Primosten, Yugoslavia, 1973. V. 1. P. 248–278.
11. Шурыгин А. М. Оценка параметров нормального распределения по “загрязненной” выборке и статистический кластер-критерий // Теория вероятностей. 1979. № 1. С. 233–234.
12. Шурыгин А. М. Оценки параметров нормального распределения с экспоненциальным взвешиванием наблюдений: асимптотическая теория // Алгоритмическое и программное обеспечение прикладного статистического анализа. М.: Наука, 1980. С. 241–259.
13. Marrona R. A. Robust M-estimators of multivariate location and scatter // Ann. Math. Statist. 1976. V. 4. № 1. P. 51–57.
14. Шурыгин А. М. Робастность и устойчивость статистических оценок // Статистика, вероятность, экономика. М.: Наука, 1985. С. 90–98.
15. Шурыгин А. М. Устойчивые оценки параметров многомерного нормального распределения // Статистика, вероятность, экономика. М.: Наука, 1985. С. 314–317.
16. Шурыгин А. М. Вариационная оптимизация устойчивости оценки // АИТ. 1994. № 11. С. 73–86.
17. Шурыгин А. М. Пути улучшения дискриминации в нормальном случае // Статистика, вероятность, экономика. М.: Наука, 1985. С. 379–382.
18. Айвазян С. А., Буштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989.
19. Шурыгин А. М. Статистический кластер-критерий // Алгоритмическое и программное обеспечение прикладного статистического анализа. М.: Наука, 1980. С. 360–368.
20. Rendall M. G., Stuart A. D. The advanced theory of statistics. V. 3: Design and analysis, and time series. London: Charles Griffin, 1968. (Перевод: Кендалл М., Стьюарт А. Многомерный статистический анализ и временные ряды. М.: Наука, 1976.)
21. Прудников А. П., Брычков Ю. А., Маричев А. И. Интегралы и ряды. М.: Наука, 1981.

Поступила в редакцию 18.01.94