

Бражников Андрей Романович

магистрант, кафедра прикладной математики и информационных технологий, ФГБОУ ВО «Донецкий государственный университет», Россия,

г. Донецк

ГАУССОВА АППРОКСИМАЦИЯ ДИВЕРГЕНЦИИ ЙЕНСЕНА-ШЕННОНА: ГРАНИЦА $\ln 2$, ЭФФЕКТ НАСЫЩЕНИЯ И КОРРЕКЦИЯ ИНТЕГРАЛЬНОГО ПОКАЗАТЕЛЯ СЛОЖНОСТИ

Аннотация. В интегральном показателе информационной сложности многомерных нестационарных систем статистическая компонента вычисляется через гауссову аппроксимацию дивергенции Йенсена–Шеннона. В расчётном эксперименте на курсах валют Банка России зафиксировано значение 1,386 нат, вдвое превышающее теоретический потолок $\ln 2 \approx 0,693$ нат. Доказано, что противоречия нет: подстановка суррогатной плотности вместо смеси распределений завышает оценку ровно на расхождение Кульбака–Лейблера между смесью и суррогатом, а для моментно-согласованной гауссианы это смещение равно негэнтропии смеси и сверху не ограничено. Публикуемая величина есть верхняя оценка дивергенции, а её избыток над $\ln 2$ измеряет негауссовость смеси. Методом Монте-Карло установлено условие насыщения: уровень $0,99\ln 2$ достигается при расстоянии Махаланобиса 5,63 независимо от размерности пространства. Предложена логарифмическая ренормировка, восстанавливающая измеримость остроты перехода: скорректированная структурная доля санкционного пика 2022 г. составляет 0,90-0,91 против опубликованных 0,81.

Ключевые слова: дивергенция Йенсена–Шеннона, гауссова аппроксимация, эффект насыщения, информационная сложность, расстояние Махаланобиса, негэнтропия, логарифмическая ренормировка, критические переходы.

Brazhnikov Andrey Romanovich

*Master's student, Department of Applied Mathematics and Information
Technology, Donetsk State University, Russia, Donetsk*

GAUSSIAN APPROXIMATION OF THE JENSEN–SHANNON DIVERGENCE: THE $\ln 2$ BOUND, THE SATURATION EFFECT AND A CORRECTION OF AN INTEGRAL COMPLEXITY INDICATOR

Annotation. In an integral indicator of the information complexity of multidimensional non-stationary systems, the statistical component is computed through a Gaussian approximation of the Jensen–Shannon divergence. A computational experiment on Bank of Russia exchange rates yielded 1.386 nats, twice the theoretical ceiling of $\ln 2 \approx 0.693$ nats. It is proved that there is no contradiction: substituting a surrogate density for the mixture of distributions inflates the estimate by exactly the Kullback–Leibler divergence between the mixture and the surrogate, and for a moment-matched Gaussian this bias equals the negentropy of the mixture and is unbounded above. The reported quantity is an upper bound on the divergence, and its excess over $\ln 2$ measures the non-Gaussianity of the mixture. Monte Carlo simulation establishes the saturation condition: the level $0.99\ln 2$ is attained at a Mahalanobis distance of 5.63 irrespective of the dimension of the space. A logarithmic renormalization is proposed that restores the measurability of transition severity: the corrected structural share of the 2022 sanction peak is 0.90–0.91 versus the published 0.81.

Keywords: Jensen–Shannon divergence, Gaussian approximation, saturation effect, information complexity, Mahalanobis distance, negentropy, logarithmic renormalization, critical transitions.

Введение

Феноменологический подход трактует сложность как измеримое отклонение состояния системы от опорного режима [1]. На этой основе в [2] построен интегральный показатель информационной сложности:

$$C_I(t) = D_{JS}(P_t \| P_{ref}) + \lambda \Delta_{struct}(t) \quad (1)$$

где $C_I(t)$ – интегральный показатель, нат; $D_{JS}(P_t \| P_{ref})$ – статистическая компонента (дивергенция Йенсена-Шеннона между текущим и опорным распределениями), нат; λ – весовой коэффициент, нат; $\Delta_{struct}(t)$ – структурная компонента (нормированное расстояние Фробениуса между ковариационными матрицами), безразмерная.

Показатель верифицирован на курсах шести инструментов Банка России за 2018-2026 гг. (окно 252 дня); для санкционного шока 2022 г. опубликовано: максимум 7,45, статистическая компонента 1,386 нат, вклад структурной 6,06, её доля 0,81 при $\lambda = 0,4068$ [2]. Но дивергенция Йенсена-Шеннона ограничена величиной $\ln 2 \approx 0,693$ нат [6], причём ограниченность есть свойство самой меры: зафиксированное значение превышает предел вдвое. Относительно того же порога вычислены и доли «насыщенных» окон: 19,4%, 25,1% и 31,3% при окне 252, 120 и 60 дней [2]. Цель работы – объяснить превышение и устранить дефект измеримости.

Смещение гауссовой оценки

Согласно [2], расхождения Кульбака–Лейблера берутся из аналитического выражения, справедливого лишь для пары нормальных распределений, тогда как смесь $M = (P + Q)/2$ гауссовой не является и потому неявно подменяется гауссовой плотностью g [5]. Записав дивергенцию через дифференциальные энтропии, $D_{JS} = H(M) - \frac{1}{2}H(P) - \frac{1}{2}H(Q)$ [4, с. 243–256], и подставив суррогат вместо смеси, получаем лемму о смещении:

$$D_{JS}^{(g)} = \frac{1}{2}D_{KL}(P \| g) + \frac{1}{2}D_{KL}(Q \| g) = D_{JS}(P \| Q) + D_{KL}(M \| g) \quad (2)$$

где $D_{JS}^{(g)}$ – оценка, полученная подстановкой суррогата g вместо смеси M , нат; $D_{KL}(\cdot \| \cdot)$ – расхождение Кульбака–Лейблера, нат.

Доказательство: полусумма интегралов $\int P \ln\left(\frac{P}{g}\right)$ и $\int Q \ln\left(\frac{Q}{g}\right)$ равна $-\frac{1}{2}H(P) - \frac{1}{2}H(Q) - \int M \ln g$; вычитая определение

дивергенции, получаем $D_{KL}(M \parallel g)$. Для моментно-согласованной гауссианы перекрёстная энтропия смеси совпадает с энтропией суррогата, откуда

$$D_{JS}^G = D_{JS} + \Delta H, \quad \Delta H = H(g) - H(M) = D_{KL}(M \parallel g) \geq 0 \quad (3)$$

где D_{JS}^G – гауссова оценка дивергенции (вычисляемая в [2] величина), нат; ΔH – смещение, равное негэнтропии смеси, нат.

Неотрицательность ΔH следует из максимальности дифференциальной энтропии нормального закона при фиксированной ковариации [4, с. 254]. Отсюда три следствия. Гауссова оценка – верхняя граница истинной дивергенции, величиной $\ln 2$ не ограниченная: 1,386 нат теории не противоречит и дивергенцией Йенсена-Шеннона не является. Избыток над потолком – нижняя оценка негауссовости смеси: для пика 2022 г. $\Delta H \geq 0,693$ нат, то есть «аномальное» число несёт информацию. Доли насыщенных окон в [2] вычислены по гауссовой оценке и потому суть верхние границы истинных.

Условие насыщения

При равных ковариациях задача редуцируется к одномерной, и дивергенция зависит только от расстояния Махаланобиса d , не завися от размерности:

$$D_{JS}(d) = \ln 2 - E_Z[\ln(1 + e^{dZ - d^2/2})] \quad (4)$$

где d – расстояние Махаланобиса между центрами распределений, безразмерное; Z – стандартная нормальная величина, $Z \sim N(0, 1)$.

Гауссова оценка в том же канале равна $\frac{1}{2} \ln(1 + \frac{d^2}{4})$ и не ограничена. Расчёт автора (Монте-Карло, $2 \cdot 10^6$ выборок, контроль квадратурами): уровень $0,99 \ln 2$ достигается при $d^* = 5,63$ одинаково для $n = 1$ и $n = 6$. За порогом индикатор слеп – при росте d с 5,63 до 7,74 истинная дивергенция меняется на 0,0004 нат, а гауссова оценка возрастает с 1,094 до 1,386 нат (рисунок 1). Реконструкция пика 2022 г. даёт истинную дивергенцию 0,572-0,693 нат (82-100% потолка): показатель работал в зоне насыщения. В

структурном канале ($\Sigma_t = k\Sigma_{ref}$) при $n = 6$ тот же уровень достигается при $k^* = 40,7$.

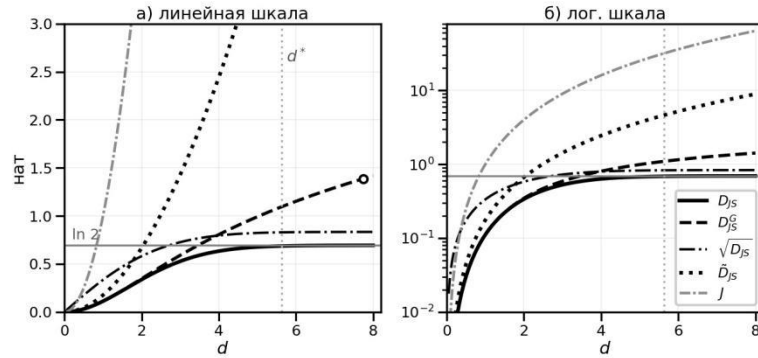


Рисунок 1. Меры расхождения в зависимости от расстояния Махаланобиса $d(n = 6)$: а) линейная шкала; б) логарифмическая. Горизонталь – потолок $\ln 2$, вертикаль – порог $d^* = 5,63$, маркер – значение 1,386 нат из [2]. Расчёт автора

Коррекция и трёхкомпонентный портрет

Корень $\sqrt{D_{JS}}$ ограничен величиной 0,833 и насыщения не снимает. Оценка методом k ближайших соседей [3] снимает гауссово допущение, но не потолок: ограниченность есть свойство самой меры, а не способа её оценивания, поэтому заявленное в [2] направление решает задачу нормальности, но не задачу насыщения. Дивергенция Джеффриса – сумма обеих форм расхождения Кульбака-Лейблера [4] – неограничена, но лишена нормировки и несопоставима между доменами. Предлагается логарифмическая ренормировка:

$$\tilde{D}_{JS} = -\ln\left(1 - \frac{D_{JS}}{\ln 2}\right) \quad (5)$$

где $\tilde{D}_{JS}$ – ренормированная статистическая компонента, безразмерная, принимает значения от нуля до бесконечности.

Преобразование строго монотонно, поэтому площадь под ROC-кривой не меняется: ренормировка не улучшает детекцию – она возвращает измеримость остроты перехода и корректность доли компонент. Модифицированный показатель $\tilde{C}_I = \tilde{D}_{JS} + \tilde{\lambda}\Delta_{struct}$; третьей компонентой портрета берётся смещение $G(t) = D_{JS}^G - D_{JS} = D_{KL}(M_t || g_t) \geq 0$, которое из погрешности

становится измеряемой величиной. Пересчёт пика 2022 г. при вкладе структурной компоненты 6,06 даёт 6,63-6,75 вместо 7,45 и структурную долю 0,897-0,914 вместо 0,81.

Заключение

Превышение потолка ошибкой не является: гауссова оценка смещена ровно на негэнтропию смеси. Истинная дивергенция насыщается при $d^* = 5,63$ и $k^* = 40,7$, из-за чего пик 2022 г. пришёлся на зону слепоты индикатора. Ренормировка не улучшает детекцию, но возвращает измеримость остроты перехода: скорректированная структурная доля пика 2022 г. – 0,90-0,91 против 0,81. Дальнейшая работа – пересчёт эмпирической базы [2] в трёхкомпонентной шкале.

Список литературы

1. Аверин Г.В., Звягинцева А.В., Бражников А.Р. Феноменологический подход к комплексной оценке сложности многомерных систем // Вестник Донецкого национального университета. Серия Г: Технические науки. – 2026. – № 1. – С. 124-140.
2. Звягинцева А.В., Бражников А.Р. Интегральный показатель информационной сложности многомерных систем на основе JS-дивергенции и структурного расстояния // Вестник Донецкого национального университета. Серия Г: Технические науки. – 2026. – № 2. – DOI 10.5281/zenodo.15586055.
3. Козаченко Л.Ф., Леоненко Н.Н. О статистической оценке энтропии случайного вектора // Проблемы передачи информации. – 1987. – Т. 23, вып. 2. – С. 9-16.
4. Cover T.M., Thomas J.A. Elements of Information Theory. – 2nd ed. – Hoboken: Wiley-Interscience, 2006. – 748 p.
5. Hershey J.R., Olsen P.A. Approximating the Kullback–Leibler divergence between Gaussian mixture models // Proc. IEEE Int. Conf. on Acoustics,

Speech and Signal Processing (ICASSP). – Honolulu, 2007. – Vol. 4. – P. 317-320.

6. Lin J. Divergence measures based on the Shannon entropy // IEEE Transactions on Information Theory. – 1991. – Vol. 37, № 1. – P. 145-151.

References

1. Averin G.V., Zvyagintseva A.V., Brazhnikov A.R. A phenomenological approach to the integrated assessment of the complexity of multidimensional systems // Vestnik of Donetsk National University. Series G: Technical Sciences. – 2026. – No. 1. – P. 124–140.
2. Zvyagintseva A.V., Brazhnikov A.R. An integral indicator of the information complexity of multidimensional systems based on the JS-divergence and the structural distance // Vestnik of Donetsk National University. Series G: Technical Sciences. – 2026. – No. 2. – DOI 10.5281/zenodo.15586055.
3. Kozachenko L.F., Leonenko N.N. Sample estimate of the entropy of a random vector // Problems of Information Transmission. – 1987. – Vol. 23, iss. 2. – P. 9–16.
4. Cover T.M., Thomas J.A. Elements of Information Theory. – 2nd ed. – Hoboken: Wiley-Interscience, 2006. – 748 p.
5. Hershey J.R., Olsen P.A. Approximating the Kullback–Leibler divergence between Gaussian mixture models // Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP). – Honolulu, 2007. – Vol. 4. – P. 317–320.
6. Lin J. Divergence measures based on the Shannon entropy // IEEE Transactions on Information Theory. – 1991. – Vol. 37, No. 1. – P. 145–151.