КИНЕМАТИКА
И ФИЗИКА
НЕБЕСНЫХ
ТЕЛ том 9 № 1 1993

УДК 521.9:528.111:519.24

И. В. Джунь

О числе градаций гистограмм ошибок астрономических наблюдений

Предложена простая формула для оценки числа градаций гистограмм при n ошибках наблюдений: $r = \sqrt{n}/2$ *(в более общем виде* $r = \sqrt{n/(2\pi e)}\,e^{1/p}$ *или* $r = 0.5[nm(m-0.5)(m+1)^{-1}(m-1.5)^{-1}]^{0.5}k_э$ *соответственно для* L_p*- и Пирсона VII типа генеральных совокупностей с показателями степени p, m;* $k_э = e^H/\sqrt{2\pi e \mu_2}$, *H — вероятностная энтропия). Показана целесообразность ее применения в астрометрии.*

ПРО КІЛЬКІСТЬ ГРАДАЦІЙ ГІСТОГРАМ ПОХИБОК АСТРОНОМІЧНИХ СПОСТЕРЕЖЕНЬ, Джунь І. В. — Запропоновано просту формулу для оцінки кількості градацій гістограм при n похибках спостережень: $r = \sqrt{n}/2$ *(в більш загальному вигляді* $r = \sqrt{n/(2\pi e)}\,e^{1/p}$ *чи* $r = 0.5[nm(m-0.5)\times(m+1)^{-1}(m-1.5)^{-1}]^{0.5}k_э$ *відповідно для* L_p*- та Пірсона VII типу генеральних сукупностей з показниками степеня p, m;* $k_э = e^H/\sqrt{2\pi e \mu_2}$, *H — ймовірнісна ентропія). Показано доцільність її застосування в астрометрії.*

ON THE NUMBER OF INTERVALS IN THE HISTOGRAMS OF ASTRONOMICAL OBSERVATION ERRORS, by Dzhun' I. V. — A simple formula is proposed for the choice of the number of histogram intervals for the number n of observation errors: $r = \sqrt{n}/2$ *(in a more general form,* $r = \sqrt{n/(2\pi e)}\,e^{1/p}$ *or* $r = 0.5[nm(m-0.5)(m+1)^{-1}(m-1.5)^{-1}]^{0.5}k_e$ *for the* L_p *and Pearson Type VII populations with the degrees p and m, respectively,* $k_e = e^H/\sqrt{2\pi e \mu_2}$, *H is the information entropy). The application of this formula in astrometry is shown to be advantageous.*

По вопросу выбора числа градаций гистограмм существует множество разнообразных правил, среди которых астрометристу часто не просто ориентироваться. Их обстоятельный обзор приведен в [5, 12]. Большинство рекомендаций изложены в известных пособиях по теории вероятностей и математической статистике, например, в [1, 10, 15, 17], и предназначены в основном для обработки технических данных, объем которых не превышает нескольких сотен измерений. Астрономические или космические данные могут содержать тысячи, а то и сотни тысяч измерений. Это обстоятельство вызывает некоторые сомнения в правомочности использования общих рекомендаций, приведенных в упомянутых руководствах. По-видимому, вопрос о числе интервалов гистограммы пока не имеет теоретического обоснования (см., например, [17, с. 294]).

Результаты обширных экспериментальных исследований, изложенные в [12, с. 179], показывают сильную зависимость числа градаций r от эксцесса распределения:

$$r = \frac{1}{3}\sqrt[5]{\beta_2^4 n^2}\,, \tag{1}$$

где $\beta_2 = \mu_4/\mu_2^2 = \varepsilon + 3$; n — число измерений.

Согласно [3] ряды ошибок астрономических наблюдений имеют эксцессы, изменяющиеся в довольно широких пределах $-0.12 \le \varepsilon \le 6.00$ (для современных наблюдений $+0.16 \le \varepsilon \le 6.00$). Вопрос о правомочности использования выражения (1) при исследовании ошибок в астрометрии остается открытым.

Поэтому мы предлагаем общий подход к решению вопроса оценки числа градаций гистограммы, основываясь на информационных методах анализа ошибок, изложенных в [4] и ранее обоснованных в [11]. При этом мы будем:

1. Использовать вероятностную энтропию H как меру неопределенности ряда негауссовых ошибок в виде e^H.

2. Полагать, что зона нечувствительности (ширина интервала) гистограммы есть функция энтропии $H_ц$ центра выборки относительно центра генеральной совокупности. Так как при нормальной совокупности [4]

$$e^{H_ц} = \sigma\sqrt{\frac{2\pi e}{n}}\,, \tag{2}$$

где $\sigma^2 = \mu_2$ — второй центральный момент, то, придерживаясь обычного в астрометрии доверительного правила в $\pm\sigma/\sqrt{n}$, будем считать $2e^{H_ц}$ энтропийной мерой зоны нечувствительности. Таким образом, сущность нашего подхода к решению вопроса оценки числа градаций гистрограммы сводится к следующему общему соотношению:

$$r = e^H/(2e^{H_ц}). \tag{3}$$

Здесь H — общая энтропия ряда ошибок, следующих тому или иному закону распределения; $H_ц$ — энтропия положения центра выборки относительно центра генеральной совокупности.

Перейдем теперь к приложениям формулы (3). Поскольку ошибкам астрономических определений присущи иногда, кроме положительных, и слабые отрицательные эксцессы, то для аппроксимации распределений погрешностей удобно использовать L_p-распределение, имеющее плотность вероятности

$$y = \frac{c_L}{\sigma_L}\, e^{-\frac{1}{p}\left|\frac{x-a_L}{\sigma_L}\right|^p}, \tag{4}$$

где $c_L = p^{1-1/p}[\,2\Gamma(1/p)\,]^{-1}$; a_L, σ_L, p — параметры распределения.

Энтропия генеральной совокупности, имеющей плотность (4), есть

$$H = -\int_{-\infty}^{\infty} y \ln y dx = \ln\left[2\sigma_L p^{1-1/p} e^{1/p}\,\Gamma(1/p)\right],$$

или

$$e^H = 2\sigma_L p^{1-1/p} e^{1/p}\,\Gamma(1/p)\,. \tag{5}$$

Найдем теперь энтропию центра выборки $H_ц$. В соответствии с [16, с. 179] медиана, как оценка центра, предпочтительней для L_p-распределения и, как известно, распределена асимптотически нормально [9] с дисперсией

$$\sigma_{\text{м}}^2 = 0.5^2[f^2(a_L)]^{-1}n^{-1}\,, \tag{6}$$

где $f(a_L) = c_L \,/\, \sigma_L$ — плотность распределения в точке a_L, т.е. уравнение (6) можно переписать в виде:

$$\sigma_{\text{м}} = \sigma_L\Gamma(1/p)\,n^{-0.5}p^{1/p-1}\,. \tag{7}$$

Тогда энтропия медианы выборки, найденная в предположении асимптотической нормальности ее распределения, есть:

$$H_{\text{ц}} = \ln(\sigma_{\text{м}}\sqrt{2\pi\,\mathrm{e}}) = \ln\left[\sigma_L\Gamma(1/p)\,(2\pi\,\mathrm{e}n^{-1})^{0.5}p^{1/p-1}\right], \tag{8}$$

или

$$\mathrm{e}^{H_{\text{ц}}} = \sigma_L\Gamma(1/p)\sqrt{\frac{2\pi\,\mathrm{e}}{n}}\,p^{1/p-1}\,. \tag{9}$$

Наконец, используя выражения (5) и (9), по формуле (3) находим число градаций гистограммы для L_p-генеральных совокупностей:

$$r = \mathrm{e}^{H}/(2\mathrm{e}^{H_{\text{ц}}}) = \sqrt{\frac{n}{2\pi\,\mathrm{e}}}\;\mathrm{e}^{1/p} \approx 0.242\sqrt{n}\;\mathrm{e}^{1/p}\,. \tag{10}$$

При практическом использовании формулы (10) параметр p (с точностью 0.5 %) можно получить по эксцессу эмпирического распределения при $-0.5 \le \varepsilon \le 3.5$, используя выражение [2]:

$$p = \frac{5\ln 5 - 6\ln 3}{\ln(\beta_2 - 2/9) + \ln\sqrt{3} - \ln 3}\,. \tag{11}$$

На основе (10) можно получить формулу числа градаций гистограммы для ряда известных распределений, например Лапласа ($p = 1$); Гаусса ($p = 2$); равномерного ($p = \infty$):

$$k_{\text{л}} = \sqrt{\frac{n}{2\pi\,\mathrm{e}}} = 0.66\sqrt{n}\;;$$

$$k_{\text{г}} = 0.40\sqrt{n}\;; \tag{12}$$

$$k_{\text{р}} = 0.24\sqrt{n}\;.$$

Результат (12) как будто подтверждает, что при одном и том же n, в зависимости от закона распределения, число градаций несколько различно. Действительно, используя уравнения (10) и (11) при вычислении r для приведенных выше крайних ε (современные астрономические наблюдения), имеем: $r_{0.16} = 0.42\sqrt{n}$; $r_6 = 0.88\sqrt{n}$, т. е. коэффициенты при $\sqrt{n}$ различаются здесь более чем в два раза. Означает ли это, что при изучении ошибок астрономических наблюдений интервал гистограммы нужно выбирать с учетом величины эксцесса распределения? На этот вопрос можно было бы ответить однозначно, если бы ошибки астрономических наблюдений в точности следовали L_p-распределению. Но ошибки современных астрономических наблюдений как будто еще лучше следуют распределению Пирсона VII типа. Это подтверждает сравнительный анализ вероятностей этих двух гипотез, полученный на основе критерия χ^2, например, для наблюденных и вычисленных разностей O—C, найденных из лазерных наблюдений ИСЗ по короткой программе MERIT [8].

В таблице мы приводим эти вероятности для шести дуг наблюдений

Вероятности гипотез о том, что разности $O-C$, найденные по лазерным наблюдениям ИСЗ по короткой программе MERIT, есть выборки из генеральных совокупностей Гаусса, L_p и Пирсона VII типа

Тип гипотезы	Наименования дуг и число наблюдений						
	66.5—71.5, 658	71.5—76.5, 575	76.5—81.5, 813	81.5—86.5, 817	86.5—91.5, 732	91.5—96.5, 880	66.5—96.5, 4475
Гаусса	0.000	0.025	0.000	0.011	1.482	0.000	0.000
L_p	0.014	12.72	0.020	0.004	1.441	0.174	0.002
Пирсона VII типа	0.000	32.35	0.055	0.005	0.901	7.359	8.002

ИСЗ в предположении, что анализируемые остатки $O-C$ есть выборки из гауссовой, L_p- и Пирсона VII типа генеральных совокупностей.

Из таблицы видно, что если исключить дуги 66.5—71.5 и 86.5—91.5, по которым для L_p-гипотезы вероятности выше, чем для распределения Пирсона VII типа, то по всем остальным дугам и по суммарному распределению дело обстоит как раз наоборот. В среднем вероятности того, что данные $O-C$ есть выборки из L_p- и Пирсона VII типа генеральных совокупностей равны соответственно 2.0 % и 8.1 %. Поэтому проверим наш вывод о зависимости числа интервалов гистограммы от формы распределения, опираясь на более вероятную гипотезу.

Плотность вероятности распределения Пирсона VII типа:

$$y = \frac{c_{\text{VII}}}{\sigma_{\text{VII}}}\left[1 + \frac{0.5}{M}\left(\frac{x - a_{\text{VII}}}{\sigma_{\text{VII}}}\right)^2\right]^{-m}, \tag{13}$$

где $c_{\text{VII}} = \Gamma\,(m+1)\,[\sqrt{2\pi(m-0.5)}\;\Gamma(m+0.5)\,]^{-1}$; $M = (m-0.5)^3 m^{-2}$.

Воспользовавшись интегралами в [13, с. 504, 595] находим энтропию распределения (13):

$$H = -\int\limits_{-\infty}^{\infty} y\ln y dx = \ln\left\{\frac{c_{\text{VII}}}{\sigma_{\text{VII}}}\,e^{m[\psi(m)-\psi(m-0.5)]}\right\}.$$

Раскрывая в этой формуле значение c_{VII}, имеем

$$e^H = \sigma_{VII}\frac{\Gamma(m+0.5)}{\Gamma(m+1)}\,[2\pi\,(m-0.5)\,]^{0.5}\,e^{m[\psi(m)-\psi(m-0.5)]}. \tag{14}$$

Для оценок центра распределения Пирсона VII типа обычно применяют метод максимального правдоподобия. Их дисперсия в этом случае соответствует нижней границе в неравенстве Рао — Крамера [6, 7]:

$$\sigma^2_{\hat{ц}} \ge \sigma^2_{\text{VII}}\,n^{-1}(m-0.5)^2(m+1)m^{-3}. \tag{15}$$

Тогда аналогично (8) можно записать

$$e^{H_ц} = \sigma_{\hat{ц}}\sqrt{2\pi e}\;. \tag{16}$$

Подставляя в (16) выражение (15), имеем

$$e^{H_ц} = \sigma_{\text{VII}}\left[\frac{(m-0.5)^2(m+1)}{m^3}\,\frac{2\pi e}{n}\right]^{0.5}. \tag{17}$$

Наконец, с учетом (14) и (17) находим по формуле (3) число интервалов гистограммы для генеральных совокупностей Пирсона VII типа:

$$r = \frac{1}{2}\left[\frac{nm(m-0.5)}{(m+1)(m-1.5)}\right]^{0.5} k_э, \tag{18}$$

где $k_э$ — энтропийный коэффициент распределения, равный $e^H/\sqrt{2\pi e \mu_2}$. Для распределения Гаусса $\varepsilon = 0\,(m = \infty)$ и распределения, по эксцессу соответствующего лапласовому $\varepsilon = 3$ $(m = 3.5)$, находим: $r_{m=\infty} = 0.50\sqrt{n}$; $r_{m=3.5} = 0.52\sqrt{n}$. Сравнивая этот результат с (12), видим, что использование регулярных семейств распределений Пирсона VII типа приводит к существенно более слабой зависимости величины r от типа закона распределения. Например, разность коэффициентов при $\sqrt{n}$ в уравнении (18) при $\varepsilon = 0$ и $\varepsilon = 3$ составляет всего 4.8 % против разности аналогичных коэффициентов в (12), равной 65 %. Даже для наименьшего из предложенных значений $m = 2$ для аппроксимации ошибок (рекомендация Хьюбера в [14]), по формуле (18) имеем $r_{m=2} = 0.58\sqrt{n}$. Для среднестатистического эксцесса ошибок астрономических наблюдений $\varepsilon = 1.32$ ($p = 1.305$, $m = 4.80$) находим, что выражения (10) и (18) приводят практически к одному и тому же результату:

$$r_{p=1.305} = \sqrt{\frac{n}{2\pi e}}\, e^{1/p} = 0.52\sqrt{n} ;$$

$$r_{m=4.80} = 0.5\,[nm(m-0.5)(m+1)(m-1.5)^{-1}]^{0.5}\, k_э = 0.51\sqrt{n} .$$

Поэтому для вычисления числа градаций гистограмм при исследовании рядов ошибок в астрометрии можно использовать (практически без уменьшения точности) следующее весьма простое и общее правило:

$$r = \sqrt{n}/2, \tag{19}$$

которое лишь фактором 0.5 отличается от рекомендованного Хайнхольдом и Гаеде [18]: $r = \sqrt{n}$, полученного иным путем. Правило (19) можно широко использовать не только в астрометрии, но и вообще в астрономии при статистическом изучении симметричных распределений параметров различных объектов с эксцессами $0.00 \le \varepsilon \le 3.00$. При отрицательных эксцессах можно руководствоваться формулой (10); при положительных эксцессах вне указанного интервала — формулой (18).

1. *Вентцель Е. С.* Теория вероятностей. — М.: Наука, 1969.—576 с.
2. *Джунь И. В.* Об одном обобщении математической формы распределений Лапласа и Гаусса и его применениии при математической обработке астрономических наблюдений // Кинематика и физика небес. тел.—1985.—1, № 4.—С. 62—66.
3. *Джунь И. В.* Некоторые аспекты практического использования L_p- и эксцесс-оценок при обработке геодезических измерений // Изв. вузов. Геодезия и аэрофотосъемка.—1986.— № 4.—С. 43—48.
4. *Джунь И. В.* Об учете эксцессов распределения ошибок при сравнении точности различных рядов астрометрических наблюдений // Кинематика и физика небес. тел.—1986.—2, № 1.—С. 88—94.
5. *Джунь И. В.* О выборе интервала гистограммы. — М., 1987.—10 с.—Рукопись деп. в ОНТИ ЦНИИГАиК, № 233-гд.Деп.
6. *Джунь И. В.* О границах неравенства Рао — Крамера для дисперсий оценок параметров распределения Пирсона VII типа // Кинематика и физика небес. тел.—1988.—4, № 1.— С. 85—87.
7. *Джунь И. В.* Об аппроксимации плотности вероятности некоторых рядов ошибок геодезических измерений распределением Пирсона VII типа // Изв. вузов. Геодезия и аэрофотосъемка.—1989.—№ 6.—С. 43—48.
8. *Джунь И. В.* Распределение Пирсона VII типа ошибок лазерных наблюдений ИСЗ // Кинематика и физика небес. тел.—1991.—7, № 3.—С. 82—91.

9. *Крамер Г.* Математические методы статистики. — М.: Мир, 1975.—648 с.
10. *Митропольский А. К.* Техника статистических вычислений — М.: Наука, 1971.—576 с.
11. *Новицкий П. В.* Основы информационной теории измерительных устройств. — Л.: Энергия, 1968.—248 с.
12. *Новицкий П. В., Зограф И. А.* Оценка погрешностей результатов измерений — Л.: Энергоатомиздат, 1961.—304 с.
13. *Прудников А. П., Брычков Ю. А., Маричев О. И.* Интегралы и ряды. — М.: Наука, 1981.—800 с.
14. *Робастность* в статистике. Подход на основе функций влияния: Пер. с англ. — М.: Мир, 1989.—512 с.
15. *Смирнов Н. В., Дунин-Барковский И. В.* Курс теории вероятностей и математической статистики для технических приложений. — М.: Наука, 1969.—512 с.
16. *Статистические* методы в экспериментальной физике / Под ред. А. А. Тяпкина. — М.: Атомиздат, 1976.—335 с.
17. *Щиголев Б. М.* Математическая обработка наблюдений. — М.: Наука, 1969.—344 с.
18. *Heinhold I., Gaede K. W.* Ingenieur Statistik. — München, Wien: Springer Verlag, 1964.—352 S.

Ровенский научно-исследовательский технологии машиностроения

Поступила в редакцию 02.04.92

РЕФЕРАТ ПРЕПРИНТА

УДК 521.97

ETALON-1, -2 CENTRE OF MASS CORRECTION AND ARRAY REFLECTIVITY: ПОПРАВКА К ЦЕНТРУ МАСС И ОТРАЖАТЕЛЬНАЯ СПОСОБНОСТЬ ИСЗ ЭТАЛОН-1, -2 / Миронов Н. Т., Емец А. И., Захаров А. Н., Чеботарев В. Е.

(*Препринт / АН Украины, Главная астрономическая обсерватория; МАО-92-11Е*)

В 1989 г. в СССР были запущены на высокую орбиту ($H - 20\,000$ км) два идентичных пассивных ИСЗ ЭТАЛОН-1, -2 (диаметр 1294 мм), оснащенных уголковыми отражателями (2146 отражателей на каждом ИСЗ), предназначенных для лазерных измерений расстояний с целью решения фундаментальных и прикладных задач. При редукции и обработке лазерных измерений расстояний до этих ИСЗ рекомендовалось использовать значение поправки к центру масс (расстояние плоскости вероятного отражения лазерного излучения от центра масс), равное 558 мм. Однако международная кампания лазерных наблюдений ИСЗ ЭТАЛОН-1, -2 (1 сентября — 1 декабря 1990 г.) и обсуждение результатов наблюдений этих ИСЗ на МЕждународном симпозиуме «ЭТАЛОН-91» (Москва, 3—9 июня 1991 г.) показали необходимость проведения исследований и экспериментов по определению и уточнению поправки к центру масс и отражательной способности при лазерных измерениях расстояний. Используя информацию о конструкции ИСЗ ЭТАЛОН-1, -2, геометрических и оптических характеристиках отражателей и их геометрическом распределении на поверхности спутника, мы исследовали вопрос о редукции измеренных топоцентрических расстояний к центру масс этих ИСЗ. Путем моделирования были вычислены значения поправки к центру масс ρ и отражательной способности E, исследованы вариации этих величин в зависимости от ориентаций ИСЗ ЭТАЛОН-1, -2 относительно направления падения на них лазерного излучения. Были рассмотрены 2522 возможных ориентации, для каждой из которых вычислены значения ρ и E. Значения поправки ρ к центру масс ИСЗ ЭТАЛОН-1, -2 варьируют от 567.3 мм до 583.4 мм, а среднее значение равно 576.0 мм со стандартным отклонением 3.2 мм. Отражательная способность ИСЗ, выраженная в эквивалентных уголковых отражателях, варьирует от 53.8 до 74.2 отражателей, а среднее значение равно 65.8 со стандартным отклонением 3.9. Обнаруженные вариации значений поправки к центру масс ρ и отражательной способности E спутников ЭТАЛОН-1, -2 объясняются особенностями симметрии в расположении уголковых отражателей на поверхности спутника.