

УДК: 519.233.3:519.234.3

АНАЛИЗ ПРИМЕНЕНИЯ ПАРАМЕТРИЧЕСКОГО И НЕПАРАМЕТРИЧЕСКОГО МЕТОДОВ ОБРАБОТКИ НА ПРИМЕРЕ ДАННЫХ ЗНАЧЕНИЙ СИЛЫ ТЯЖЕСТИ

Наталья Николаевна Кобелева

Сибирский государственный университет геосистем и технологий, 630108, Россия, г. Новосибирск, ул. Плеханова, 10, кандидат технических наук, доцент кафедры космической и физической геодезии, тел. (383)361-01-59, e-mail: n.n.kobeleva@mail.ru

Зарина Баранкуловна Бексултанова

Коммунальное государственное казенное предприятие «Высший колледж геодезии и картографии» управления образования Восточно-Казахстанской области, 071400, Казахстан, г. Семей, ул. Заря 42, бакалавр геодезии и картографии, преподаватель специальных дисциплин, тел. 8 -778 -470-74-62, e-mail: beksultanova_zarina90@mail.ru

В статье рассматриваются результаты исследования применения непараметрической обработки геодезических измерений. Выполнен сравнительный анализ параметрической и непараметрической оценки к результатам геодезических измерений на примере определения силы тяжести. Для реализации данного исследования использовалось программное обеспечение Statgraphics 18 и мастер диаграмм Microsoft Excel. По результатам измерений вычислены такие характеристики как медиана, среднее арифметическое, стандартное отклонение и дисперсия, а так же получены значения доверительных интервалов для измерений силы тяжести в расчете двухсот семидесяти отсчетов с сейсмозащитой и без сейсмозащиты. Приведены разности между полученной шириной доверительных интервалов для двух приемов, построены диаграммы, проведен анализ полученных результатов.

Ключевые слова: непараметрический и параметрический анализ, доверительные интервалы, медиана, среднее арифметическое, дисперсия, робастные методы, выборка, гистограмма, стандартное отклонение, сила тяжести, сейсмозащита, нормальный закон распределения, зашумление, состоятельность, несмещённость и эффективность.

ANALYSIS OF THE APPLICATION OF PARAMETRIC AND NONPARAMETRIC PROCESSING METHODS ON THE EXAMPLE OF THESE VALUES OF GRAVITY

Natalia N. Kobeleva

Siberian State University of Geosystems and Technologies, 10, Plakhotnogo St., Novosibirsk, 630108, Russia, Ph. D., Associate Professor, Department of Space and Physical Geodesy, phone: (383)361-01-59, e-mail: n.n.kobeleva@mail.ru

Zarina B. Beksultanova

State utility company "Higher College of Geodesy and Cartography" of the Education Department of the East Kazakhstan region, 071400, Kazakhstan, Semey, st. Zarya 42, bachelor of geodesy and cartography, teacher of special disciplines, tel. 8 -778 -470-74-62, e-mail: beksultanova_zarina90@mail.ru

The article discusses the results of studying the application of nonparametric processing of geodetic measurements. A comparative analysis of parametric and nonparametric estimates to the results of geodetic measurements is carried out using the example of determining the force of gravity. To implement this study, Statgraphics 18 software and Microsoft Excel chart wizard were used. According to the measurement results, such characteristics as the median, arithmetic mean, standard deviation and variance were calculated, as well as the values of the confidence intervals for

gravity measurements in the calculation of two hundred and seventy readings with and without seismic protection were obtained. Differences between the obtained width of the confidence intervals for the two methods are given, diagrams are constructed, and the analysis of the results obtained is carried out.

Key words: nonparametric and parametric analysis, confidence intervals, median, arithmetic mean, variance, robust methods, sample, histogram, standard deviation, gravity, seismic protection, normal distribution, noise, consistency, unbiasedness and efficiency, observations.

Введение

Методы обработки как экспериментальных, так и натуральных данных направлены, прежде всего, на анализ, оценку и на уменьшение влияния погрешностей результатов измерений [1]. В связи с этим, на сегодняшний день статистические методы обработки данных являются достаточно разработанными, распространенными и широко применяются на практике. При этом они подразделяются на три основные группы: классические (традиционные), робастные и непараметрические.

Классические методы – это параметрические методы, построенные на гипотезе нормального распределения исследуемых данных, анализе и оценки этих параметров распределения [2].

Робастные методы ориентированы на нахождении выбросов из всей совокупности измерений, уменьшения их воздействия или удаления их из выборки [3].

Непараметрические методы - это методы которые не зависят от каких-либо параметров, и дают возможность обрабатывать результаты наблюдений «невысокого качества», о распределении которых не так много информации [4].

Классические и робастные методы обработки данных, как правило, чаще всего используют в качестве оценки центра распределения – среднее арифметическое результатов наблюдений. Робастные методы более устойчивы к выбросам и шумам, чем классические, но все же они остаются параметрическими. В отличие от них непараметрические методы не связаны с какими-либо определенными параметрами, а в качестве центра распределения принимают медиану, а не математическое ожидание искомой выборки.

Исходные данные

Ускорения силы тяжести были получены по договоренности с геодезической экспедицией №145, города Новосибирск. Наблюдения выполнялись двумя приемами динамическим методом [5]: в первом приеме на приборе ГБЛ (гравиметр баллистический лазерный) была установлена сейсмозащита, исключаяющая влияние на отсчеты внешних воздействий, во втором приеме сейсмозащита отсутствовала. Полученные значения силы тяжести получены в условной системе, размерность измерения – мили Галлы.

В выполненных приемах, производили по три цикла наблюдений, по девятисто отсчетов в каждом цикле. Измерения производились в соответствии с требованиями инструкции [6].

Исследования

В астрономо-геодезических наблюдениях, выполняя многократные измерения одной и той же величины, при одинаковых условиях, полученные результаты отличаются друг от друга на некоторую величину, то есть поддаются случайному рассеиванию [7].

Колебание полученных отсчетов ускорения силы тяжести, как массовые повторные измерения, поддаются математической закономерности распределения [8]. Для прослеживания распределения полученных результатов силы тяжести были построены гистограммы, для двух приемов (рис.1 и рис. 2).

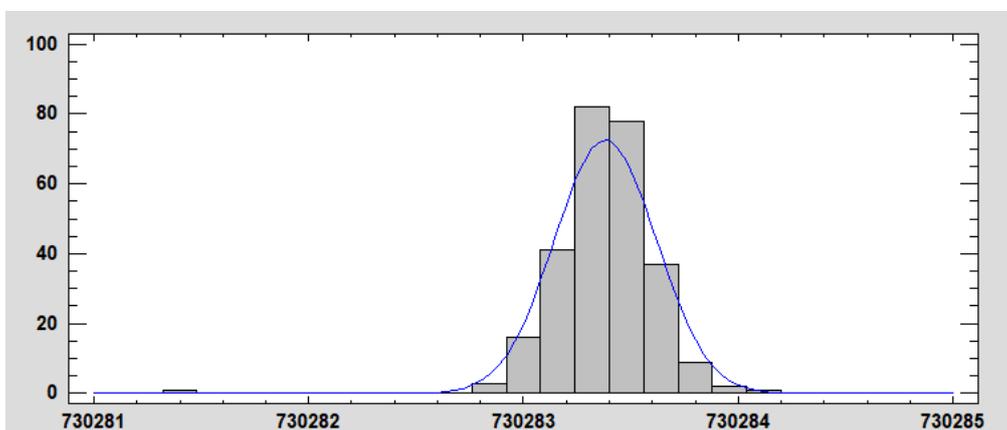


Рис. 1 Гистограмма отсчетов с сейсмозащитой

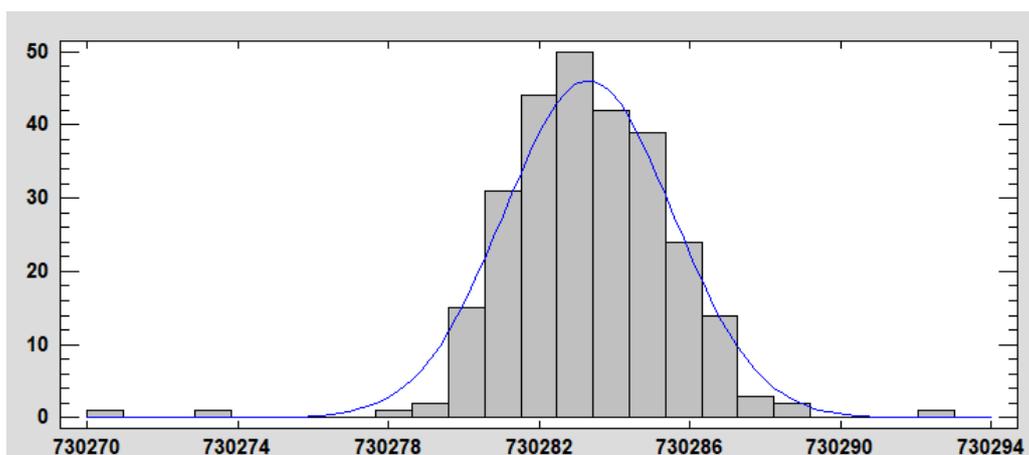


Рис.2 Гистограмма отсчетов без сейсмозащиты

Можно отметить, что с увеличением количества измерений, мы наблюдаем размещение полученных результатов ускорений силы тяжести, соответствующие нормальному закону распределения [7].

Обработку полученных ускорений силы тяжести будем выполнять используя оценки математической статистики [9]. Суть исследования будет состоять в нахождении параметрической и непараметрической оценки полученных результатов измерения силы тяжести и сравнения полученной точности.

В таблицах 1 и 2 приведены среднеарифметическая и медианная оценки измерений силы тяжести, сравнение выполнялось по всем трем циклам каждого приема: с сейсмозащитой и без.

Для анализа рассматриваемых оценок были найдены такие характеристики, как среднее арифметическое, медиана, стандартное отклонение и дисперсия [10–13].

Таблица 1

Статистические показатели оценок значений силы тяжести, полученных с сейсмозащитой

Характеристики:	1 серия	2 серия	3 серия
Среднее арифметическое	730283,4038	730283,3884	730283,2131
Медиана	730283,4170	730283,3533	730283,3335
Дисперсия по среднему арифметическому	0,03923	0,05479	0,18485
Дисперсия по медиане	0,03897	0,05541	0,19730
Стандартное отклонение по среднему арифметическому	0,19806	0,23406	0,42995
Стандартное отклонение по медиане	0,19740	0,23539	0,44418
Границы доверительного интервала по среднему арифметическому	730283,0156	730282,9296	730282,3704
	730283,7920	730283,8471	730284,0558
Границы доверительного интервала по медиане	730283,0301	730282,8919	730282,4629
	730283,8038	730283,8146	730284,2041
Границы доверительного интервала по распределению Стьюдента для среднего арифметического	730283,0102	730282,9233	730282,3588
	730283,7973	730283,8534	730284,0674
Границы доверительного интервала по распределению Стьюдента для медианы	730283,0247	730282,8855	730282,4509
	730283,8092	730283,8210	730284,2161

Вычисленные числовые характеристики (табл.1, табл.2) не дают полной информации о точности и надежности полученной оценки [14]. Показатель

точечного оценивания устанавливается такими характеристиками как состоятельность, несмещённость и эффективность [15].

Несмещенность оценки означает, что ее математическое ожидание равно значению оцениваемого параметра генеральной совокупности [16]. Для сравнения меры эффективности e точечного оценивания пользуются следующим соотношением:

$$e = \frac{M \times (\theta_1 - \theta)^2}{M \times (\theta_2 - \theta)^2}, \quad (1)$$

где θ_1 и θ_2 – оценки, характеризующие параметр θ , M - математическое ожидание.

При $e > 1$, оценка θ_2 считается более эффективнее, чем θ_1 (и наоборот), потому что она располагает меньшим рассеиванием.

Таблица 2

Статистические показатели оценок значений силы тяжести,
полученных без сейсмозащиты

Характеристики:	1 серия	2 серия	3 серия
Среднее арифметическое	730283,5793	730283,2921	730283,0226
Медиана	730283,3891	730283,2545	730282,9946
Дисперсия по среднему арифметическому	4,25768	4,91534	5,52704
Дисперсия по медиане	4,24656	4,86214	5,46642
Стандартное отклонение по среднему арифметическому	2,06342	2,21706	2,35097
Стандартное отклонение по медиане	2,06072	2,20503	2,33804
Границы доверительного интервала по среднему арифметическому	730279,5350	730278,9467	730278,4147
	730287,6236	730287,6376	730287,6305
Границы доверительного интервала по медиане	730279,3500	730278,9326	730278,4120
	730287,4281	730287,5763	730287,5771
Границы доверительного интервала по распределению Стьюдента для среднего арифметического	730279,4793	730278,8869	730278,3512
	730287,6792	730287,6974	730287,6939
Границы доверительного интервала по распределению Стьюдента для медианы	730279,2944	730278,8731	730278,3489
	730287,4837	730287,6358	730287,6402

Эффективность оценки так же можно определить соотношением дисперсий, если θ_1 и θ_2 – несмещенные оценки:

$$e = \frac{\sigma^2 \theta_1}{\sigma^2 \theta_2}, \quad (2)$$

а $\sigma^2 \theta_1$ и $\sigma^2 \theta_2$ – дисперсии соответствующих оценок.

Из формулы (2) следует, чем меньше дисперсия тем эффективнее оценка.

Состоятельность оценки определяется постоянством несмещенности оценки, при безграничных объемах выборки [7].

По результатам исследований в таблице 2 можно заметить, что значение дисперсии при непараметрическом оценивании является минимальным, исходя из этого можно сделать выводы о том, что выполняется требование несмещенности и эффективности оценки [17].

По - мимо точечного оценивания используется и интервальное. Любая статистическая оценка параметра, выполненная по полученным данным наблюдений, становится объективнее при определении границ допустимых погрешностей [7].

При сжатом числе наблюдений оценку точности рекомендовано совершать посредством нахождения доверительных интервалов [18,19].

Построение доверительных интервалов выполнялось по средней арифметической и медианой оценкам, по формулам (3) и (4) [13].

Доверительный интервал для математического ожидания при известной дисперсии σ :

$$\bar{x} - \varepsilon_\beta \times \left(\frac{\sigma}{\sqrt{n}} \right) < \alpha < \bar{x} + \varepsilon_\beta \times \left(\frac{\sigma}{\sqrt{n}} \right), \quad (3)$$

где n – объем выборки, ε_β - значение аргумента функции Лапласа, \bar{x} - выборочное среднее.

И если не известна дисперсия, то используется следующая формула:

$$\bar{x} - t_\beta \times \left(\frac{s}{\sqrt{n}} \right) < \alpha < \bar{x} + t_\beta \times \left(\frac{s}{\sqrt{n}} \right), \quad (4)$$

где s – исправленное среднее квадратические отклонение, t_β - значение аргумента функции распределения Стьюдента [20].

Анализируя результаты вычислений (табл.1, табл.2), можно сделать вывод, что если в серии наблюдений присутствуют выбросы (шум), то медианная оценка дает более точный результат. Об этом можно судить и по границам доверительных интервалов, чем они уже, тем точнее оценка. Сравнение значений границ доверительных интервалов приведены в таблице 3.

Таблица 3

Сравнение значений ширины доверительных интервалов двух приемов

Характер нахождения доверительного интервала	Ширина доверительных интервалов с сейсмозащитой			Ширина доверительных интервалов без сейсмозащиты		
	1	2	3	1	2	3
По среднему арифметическому	0,7764	0,9175	1,6854	8,0886	8,6909	9,2158
По медиане	0,7738	0,9227	1,7412	8,0780	8,6437	9,1651
Разница доверительных интервалов	-0,0026	+0,0052	+0,0561	-0,0106	-0,472	-0,0507
По распределению Стьюдента для среднего арифметического	0,7871	0,9302	1,7086	8,1999	8,8105	9,3426
По распределению Стьюдента для медианы	0,7844	0,9354	1,7652	8,1892	8,7627	9,2913
Разница доверительных интервалов	-0,0027	+0,0052	+0,0566	-0,0107	-0,0478	-0,0513

Заключение

Выполняя анализ полученных исследований можно сделать следующие первоначальные выводы о применимости непараметрического метода к математической обработке геодезических данных:

1. по данным таблицы 2, видно, что при точечном оценивании полученных наблюдений, выполняются такие требования как несмещенность и эффективность, о чем говорит постоянство минимальной дисперсии, если в качестве центра распределения принимать медиану данного вариационного ряда, а не среднеарифметическое значение;

2. при интервальной оценке, ширина доверительных интервалов уже по медианному оцениванию при наблюдениях, содержащих выбросы, т.е. в наблюдениях без сейсмозащиты. Это говорит о том, что среднеарифметическое оценивание лучше при наблюдениях, не содержащих выбросы.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Нефедова Г.А., Ащеулов В.А. Теория вероятности и математическая статистика // учеб. пособие. – Новосибирск: СГГА, 2005. – 94 с.
2. Герасимов А.Н., Морозова Н.И. Параметрические и непараметрические методы в медицинской статистике // Эпидемиология и вакцинопрофилактика №5 (84) /2015. – С. 6-12.
3. Смирнов П.О. Робастные методы и алгоритмы оценивания корреляционных характеристик данных на основе новых высокоэффективных и быстрых робастных оценок масштаба // автореферат Санкт –Петербург – 2013. – С. 18.
4. Кобелева Н.Н. Разработка информационной технологии непараметрической обработки геодезических данных // Вестник Сибирской государственной геодезической академии. – Новосибирск, 2002. – Вып. 7. – С. 139-142.

5. Шароглазова Г.А. Гравиметрия // учеб. -метод. комплекс – Новополюцк: ПГУ, 2006 – 196 с.
6. Инструкция по развитию высокоточной государственной гравиметрической сети России // М., ЦНИИГАиК 2001 66 с.
7. Смирнов Н.В., Белугин Д.А. Теория вероятностей и математическая статистика в приложении к геодезии, монография // М., «Недра», 1969, стр. 379.
8. Ламанова Н.Г. Статистическое оценивание результатов и погрешностей измерений при многократных наблюдениях // Перм.нац. исслед. политехн. ун-т. Пермь, 2017. 39с.
9. Луценко А.И. Математическая статистика // Ростов н/Д: Изд-во ЮФУ, 2014, 157с.
10. Балаш О.С. Методические указания и варианты контрольных заданий по статистике (часть 1) // Саратов – 46 с.
11. Гребенникова И.В. Методы математической обработки экспериментальных данных // уч.-методическое пособие Екатеринбург: Изд-во Урал. ун-та, 2015. — 124 с.
12. Спирип Н.А., Лавров В.В. Планирование эксперимента //Екатеринбург: Изд-во Урал. ун-та, 2015. — 124 с.
13. Лапко А. В., Лапко В. А. Непараметрические модели и алгоритмы обработки информации // учеб. пособие. - Сиб. гос. аэрокосмич. ун-т. – Красноярск, 2010. – 220 с.
14. Орлов А. И. Непараметрическое оценивание характеристик распределений вероятностей // Политематический сетевой эл. науч. журнал КГАУ. - Вып. -2015.- С. 1-20.
15. Воскобойников Ю.Е., Тимошенко Е.И. Математическая статистика (с примерами в Excel) // уч. пособие. Новосибирск: НГАСУ (Сибстрин), 2006. – 152 с.
16. Газарян А.Ш., Мокриков С.С., Точечные оценки параметров распределения // Междунар. студ. науч. Вестник - Вып.3 (часть 1) – 2018 - С. 56-58.
17. Амридон Г.Б., Барлиани И.Я. Оценка неравномерно измеренных пространственных данных, полученных методом псевдонормальной оптимизации, и их свойства // Вестник-СГУГиТ. -Том-22-№-4 2017 год с – 27-39.
18. Куштин И.Ф. Геодезия: обработка результатов измерений // учеб. пособие. – Ростов на Дону - 2006. С – 288.
19. Радионова М.В. Построение оптимальных доверительных интервалов для параметров положения и масштаба распределений // ГОУ ВПО – Пермский государственный университет, г. Пермь. - 2010. - С-7
20. Попов В. А., Бренерман М. Х. Руководство к решению задач по теории вероятностей и математической статистике // Издательство Казанского государственного университета. – 2008. - С-119.

© Н. Н. Кобелева, З. Б. Бексултанова, 2020