УДК 528.48

АНАЛИЗ И ОЦЕНКА ОДНОРОДНОСТИ ИНЖЕНЕРНО-ГЕОДЕЗИЧЕСКИХ ИЗМЕРЕНИЙ

THE ANALYSIS AND ESTIMATION OF UNIFORMITY OF ENGINEERING-GEODETIC MEASUREMENTS

Угненко Е.Б., д.т.н., проф. (Харьковский национальный автомобильно-дорожный университет, г. Харьков)

Ugnenko Ye. B., Doctor of Engineering, Professor (Kharkiv National Automobile and Highway University, Kharkiv)

В работе проведен анализ и оценка однородности инженерно-геодезических измерений. Алгоритм оценки неоднородности результатов измерения основывается на вероятностных моделях, устанавливающих связь между структурой измерения и ее отображением в пространстве параметров.

In article the analysis and an estimation of uniformity of engineering-geodetic measurements is carried out. The algorithm of an estimation of heterogeneity of results of measurement is based on the likelihood models establishing connection between structure of measurement and its display in space of parametres.

Ключевые слова: геодезические измерения, оценки погрешностей, вероятностные параметры, критерий однородности

Keywords: geodetic measurements, estimations of errors, likelihood parametres, criterion of uniformity

Одним из основных требований к точности геодезических измерений является обеспечение достоверной оценки погрешностей, сопутствующих измерительному процессу. Наиболее достоверной информацией о влиянии погрешностей измерений могут дать обширные эксперименты, выполненные при различных условиях измерения. Однако, как правило, многочисленность экспериментов реализовать не удается, в связи с чем, приходится принимать решение при неполной информации, т. е. в условиях неопределенности. Некорректная оценка погрешностей может привести к нерациональному использованию измерительной техники, необоснованному назначению технических допусков, к научно-техническим просчетам, к неправильному применению аппарата математической обработки измерений.

В последнее время в научной литературе и метрологической практике укоренилось двухуровневое направление исследования в этой области: теоретическое и натурное. В результате теоретических исследований с использованием статистического метода, могут быть найдены статистические значения вероятностных характеристик модели будущего измерения и аналитическим путем – вероятностные параметры упрощенной модели измерения. Натурная апробация выдвинутой гипотезы позволит уточнить вероятностные характеристики и принять наиболее надежное решение в отношении траектории измерений.

Грубую погрешность измерений следует рассматривать с двояких позиций: с одной стороны, как вероятностную, с другой стороны, как аномальную. Ясно, что и подход к их анализу должен быть различным.

Вероятностная ошибка не является грубой. Она не согласуется уже с основной выборкой, но еще не принадлежит другой генеральной совокупности. Вероятностная ошибка указывает, что в измерении наметился дестабилизирующий фактор, и траектория измерения имеет тенденцию выхода из настроечного уровня. Вероятностные результаты могут обрабатываться совместно с основной группой измерений специальными приемами (например, по принципу смешанных совокупностей).

Аномальная ошибка может возникнуть в результате резкой рассогласованности условий геодезических измерений. Она не отражает динамику запроектированного измерения, является образом совершенно другого фазового пространства, и результаты, геодезических измерений не могут совместно обрабатываться с основной группой измерений. Такие ошибки подлежат отбраковке.

В любом случае, анализируя природу грубых погрешностей, экспериментатор решает задачу об однородности собранной информации. Алгоритм оценки неоднородности результатов измерения основывается на вероятностных моделях, устанавливающих связь между структурой измерения и ее отображением в пространстве параметров.

С практической точки зрения задача оценки однородности двух выборочных распределений состоит в определении на основе эмпирических данных действительной вероятностной меры из семейства возможных вероятностных мер.

На этом основывается дальнейшее определение различимости вводимых вероятностных мер [1].

В качестве меры различимости вводится величина

$$\rho_1\left(P_1P_2\right)=\int_{-\infty}^{\infty}\left|f_1\left(x\right)-f_2\left(x\right)\right|dx, \tag{1}$$

где P_1 и P_2 – вероятностные меры, принадлежащие одному множеству X;

$f_1\left(x\right)$ и $f_2\left(x\right)$ – плотности распределения, характеризующие эти меры.

Расстояние $\rho_1\left(P_1P_2\right)=l\left(x\right)$ обращается в ноль, если плотности выборочных распределений совпадают, и принимает свое максимальное значение, равное 2, если плотности не пересекаются. Таким образом, $l\left(x\right)$ есть полная вариация меры $\left(P_1-P_2\right)$.

Для удобства аналитических преобразований введем видоизмененную норму $\left(\sqrt{f_1\left(x\right)}-\sqrt{f_2\left(x\right)}\right)$ в пространстве $l\left(x\right)$.

Тогда расстояние между вероятностными мерами определяется как

$$\rho_2^2\left(P_1P_2\right)=\int_{-\infty}^{\infty}\left(\sqrt{f_1\left(x\right)}-\sqrt{f_2\left(x\right)}\right)^2dx=2-2\int\sqrt{f_1\mathrm{T}_2}dx. \tag{2}$$

Величина $\rho_2\left(P_1P_2\right)$ может принимать значения в интервале $\left[0,\sqrt{2}\right]$, причем $\rho_2=0$ при $P_1=P_2$ и $\rho_2=\sqrt{2}$, когда $f_1\left(x\right)\times f_2\left(x\right)=0$, что соответствует двум непересекающимся множествам $P_1\left(A_1\right)=P_2\left(A_2\right)=1$.

Параметр ρ не зависит от выбора величины x, ибо расстояние сохраняется вне зависимости от выбора доминирующего значения аргумента.

Для любого x – измеряемого множества A справедливы, неравенства

$$\left\{\int_A \sqrt{f_1 f_2}dx\right\}^2 \le \int_A f_1 dx \int_A f_2 dx; \qquad (3)$$
$$\int_A \sqrt{f_1 f_2}dx \le \sqrt{P_1(A)P_2(A)}.$$

Из неравенств (3) следует, что если существует разбиение множества X на конечное число непересекающихся множеств $A_1,\ldots,A_n$, т.е. $X\bigcup_1^n A_2$ то

$$\int_A \sqrt{f_1 f_2}dx \le \sum_{r=1}^{n} \sqrt{P_1(A_2)P_2(A_r)}.$$

Если $\Phi_{1n}(x_1,\ldots,x_n) = f_1(x_1) f_1(x_2)\ldots f_1(x_n)$,
$\Phi_{2n}(x_1,\ldots,x_n) = f_2(x_1) f_2(x_2)\ldots f_2(x_n)$, то

$$1-\frac{1}{2}\rho_2^2(\Phi_{1n}\Phi_{2n})\int\sqrt{\Phi_{1n}\Phi_{2n}}x(dx_1)\ldots x_n(dx_n) =$$

$$=\left\{\int\sqrt{f_1 f_2 x(dx_1)}\right\}^n = \left\{1-\frac{1}{2\rho_2^2(f_1 f_2)}\right\}^n$$

и, следовательно, стремится к нулю при $n\to\infty$, если $f_1 \ne f_2$.

Таким образом, при $n\to\infty$ функция $\rho_2^2(\Phi_{1n},\Phi_{2n})\to 2$. Это значит, что если две функции эмпирического распределения различны, то при достаточно большом n можно добиться того, что величины ρ_2^2 или ρ близки к максимальному значению настолько, насколько мы этого захотим. Изложенные соображения можно применить к оценке значимости величины ρ между двумя дискретными эмпирическими распределениями [1].

Имеется генеральная совокупность объемом n с функцией распределения вероятностей P, в которой каждому значению случайной величины x соответствует вероятность P_k. Случайным образом берется выборка объемом n, с распределением вероятностей $P' = \{P_1'P_2',\ldots,P_k'\}$, где $r=\overline{1,k}$ и $P_r' = n_1/n$. Вводится в рассмотрение величина

$$X_n^2 = \sum_1^k \frac{\left(n_r - nP_r\right)^2}{nP_r} = n\sum_{r=1}^k \frac{\left(P_r' - P_r\right)^2}{P_r}. \tag{4}$$

При $n = \infty$ величина (4) по вероятности сходится к распределению χ^2 с $k-1$ степенями свободы.

В соответствии с выражением (2) будем иметь

$$\rho_2^2\left(P, P'\right) = \sum_{r=1}^k \left(\sqrt{P'}_r - \sqrt{P_r}\right)^2 = \sum_{r=1}^k \frac{\left(P_r' - P_r\right)^2}{\left(\sqrt{P_r' + \sqrt{P_r}}\right)^2} =$$

$$= \frac{1}{4}\sum_{r=1}^k \frac{\left(P_r' - P_r\right)^2}{P_r}\left[1 - \frac{\left(\sqrt{P_r' + \sqrt{P_r}}\right)\left(\sqrt{P'}_r + 3\sqrt{P_r}\right)}{\left(\sqrt{P'}_r + \sqrt{P_r}\right)^2}\right]. \tag{5}$$

Несложные преобразования приводят к выражению

$$4n\rho^2 = n\sum_{r=1}^k \frac{\left(P_r' - P_r\right)^2}{P_r}\left(1 - \varepsilon_r\right), \tag{6}$$

где $\left|\varepsilon_r\right| \le \dfrac{3\left|\sqrt{P_r'} - \sqrt{P_r}\right|}{\sqrt{P_r'} + \sqrt{P_r}} \to 0$ при $P_r' \to P_r$.

Ясно, что при $n = \infty$, $\varepsilon_r = 0 \subset P = 1$.

Рассмотрим две выборки с параметрами $P_{1n} = \dfrac{n_{1r}}{n_1}$, $P_{2n} = \dfrac{n_{2r}}{n_2}$.

Мерой различимости двух выборочных совокупностей является величина

$$X_{n_1 n_2}^2 = n_1 n_2 \sum_{r=1}^k \frac{\left(n_{1r}/n_1 - n_{2r}/n_2\right)}{n_{1r} + n_{2r}},$$

имеющая распределение χ^2.

Квадрат расстояния между этими выборками определим из выражения

$$\rho^2 = \sum_{r=1}^{k}\left(\sqrt{P_{1r}} - \sqrt{P_{2r}}\right)^2 = \frac{n_1 + n_2}{4}\sum_{r=1}^{k}\frac{\left(P_{1r} - P_{2r}\right)^2}{n_1 P_{1r} + n_2 P_{2r}}\left(1 - \varepsilon_r\right), \quad (7)$$

При $n_1 = n_2 = n$ получаем для $X^2_{n,n}$ выражение, аналогичное (4), а расстояние между выборочными совокупностями вычисляется по формуле

$$\rho^2 = \frac{1}{2}\sum_{r=1}^{k}\frac{\left(P_{1r} - P_{2r}\right)^2}{P_{1r} + P_{2r}}\left[1 + \frac{\left(\sqrt{P_{1r}} - \sqrt{P_{2r}}\right)^2}{\left(\sqrt{P_{1r}} + \sqrt{P_{2r}}\right)^2}\right]. \quad (8)$$

Очевидно, что $X^2_{n,n} \le 2n\rho^2 \le 2X^2_{n,n}$.

Практическое применение изложенных соображений заключается в вычислении $\rho^2_{\text{эмп}}$ по формуле (8). Из таблиц распределения χ^2 выбирается значение $\chi^2_{\text{табл}}$ которое сравнивается с $\rho^2_{\text{эмп}}$. Если $\rho^2_{\text{эмп}} > \chi^2_{\text{табл}}$ делается вывод о существенности расхождений выборочных совокупностей, а следовательно, о неоднородности смешиваемых групп.

Для проверки теоретических положений анализируется совокупность невязок треугольников триангуляции 2-го класса объемом 170 единиц, из которой образовывались выборки примерно равных объемов (по 29 невязок, примерно такое число приходится на одну трапецию масштаба 1:100000). Было проанализировано 15 различных комбинаций триангуляции 2-го класса и 6 комбинаций триангуляции 1-го класса с триангуляцией 2-го класса. По общим правилам статистических вычислений выборка разбивалась на 16 интервалов, строилась гистограмма эмпирического распределения, после чего подсчитывалась частота попадания невязки определенной величины в фиксированный интервал. Частоты принимались за эмпирические вероятности, которые подставлялись в формулу (8). По таблицам χ^2 распределения по принятой доверительной вероятности выбирают $\chi^2_{\text{табл}}$ и, если $\rho^2 < \chi^2_{\text{табл}}$, гипотеза об однородности анализируемых выборочных совокупностей принималась, а в противном случае –

отвергалась.

Проверка однородности двух центров выборочных распределений по критерию Стьюдента:

$$t = \frac{v_1 - v_2}{\sqrt{n\left(\sigma_1^2 \sigma_2^2\right)}} \sqrt{\frac{n_1 n_2 \left(n_1 + n_2 - 2\right)}{n_1 + n}} . \qquad (9)$$

Так как $t_{\text{расч}} < t_{\text{табл}}$, следует вывод об однородности выборочных центров распределения.

Проверка однородности выборочных дисперсий по критерию Фишера:

$$F = \frac{m_1^2}{m_2^2} . \qquad (10)$$

Так как $F_{\text{расч}} < F_{\text{табл}}$ то следует вывод об однородности двух выборочных дисперсий. Таким образом, критерий эмпирического расстояния дополняет арсенал вероятностных методов анализа выборочных совокупностей и по своей эффективности не уступает классическим критериям.

Литература

1. Сухов А.Н. Системный анализ геодезических измерений. /А. Н. Сухов/ – М.: Недра, 1991. – 325с.