Обработка многократно измеренных величин

Необходимость альтернативных оценок:

– число измерений небольшое, закон распределения неизвестен;

–распределение результатов измерений нормальное, но с засоренностью ε другим, не известным законом;

– результаты измерений достаточно нормальны, но имеют значимые отклонения по основным характеристикам случайности, и/или однородности и/или независимости.

Действия по учету нарушения характеристик;

Действия при неизвестном, или засоренном законе.

Обработка многократно измеренных величин

Доказано:

- закон распределения результатов измерений нормальный - наилучшие оценки для сдвига, СА, для масштаба СО σ;

- результаты измерений имеют закон распределения Лапласа - наилучшие оценки для сдвига, медиана, для масштаба САО;

- нормальный закон имеет небольшое загрязнение частью другого - наилучшие оценки для сдвига,

медиана, для масштаба скорректированное (АМО(x)).

Основа – метод максимального правдоподобия. 3 случай

Обработка многократно измеренных величин

Основные виды оценок в геодезии (начиная с 18 века):

- M-оценки – робастные оценки, полученные на основе метода максимального правдоподобия на основе функции правдоподобия;
- R-оценки – робастные оценки, полученные на основе ранговых критериев (порядковых статистик);
- L-оценки – робастные оценки, полученные на основе линейных комбинаций результатов измерений.

Есть другие D, W, A-оценки и другие. В геодезии используются редко.

Обработка многократно измеренных величин

Наиболее части встречаются (в том числе в геодезии)

– общие корректирующие оценки, такие как ВМНК, ОМНК или Lp-оценки;

– для *сдвига*, *усеченное среднее*, *медиана*, *псевдомедианы* и *адаптивная оценка Хогга*;

– для оценок *масштаба*, *среднее абсолютное отклонение* (*САО*), *абсолютное медианное отклонение (АМО), интерквартильный размах IQR*, *линейная оценка Даунтона*.

Альтернативные оценки – учет разных отклонений и более устойчивые оценки.

Обработка многократно измеренных величин

Важно:

- альтернативные оценки сдвига не отличаются (мало отличаются) от среднего для НЗР результатов

- альтернативные оценки сдвига значимо отличаются от среднего при нарушении основных предположений.

Вычисляют среднее и альтернативы, сравнивают, выводы. Совпадают – среднее, не совпадают – альтернативная.

Для многих основных альтернативных оценок масштаба есть коэффициент пересчета их в традиционное стандартное отклонение. Далее сравнение как в П.1

Обработка многократно измеренных величин

Общие корректирующие оценки:

1. Модификации МНК: ВМНК

$$\overline{x} = \frac{e^T \cdot W \cdot x}{e^T \cdot W \cdot e}$$

$$\sigma_i^2 = \sigma_0^2 \cdot i$$

$$W = \begin{bmatrix} \frac{1}{\sqrt{x_1}} & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & \frac{1}{\sqrt{x_n}} \end{bmatrix}, \quad W = \begin{bmatrix} \frac{1}{x_1} & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & \frac{1}{x_n} \end{bmatrix}$$

$$W = \frac{1}{\sigma_0^2} \cdot \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & n \end{bmatrix}$$

$$\rho_{v_i, v_{i-1}} \approx 1 - \frac{DW}{2}$$

Обработка многократно измеренных величин

2. Lp-оценки. Основа ММП. Основные показатели

1 – метод наименьших модулей,

2 – метод наименьших квадратов,

∞ - метод минимакса

Вычисляют по исследованиям показатель:

$$\alpha = \frac{6}{E'}$$

Взвешивают весом через показатель и остаток

$$w_i = \sigma_0^{\alpha} \cdot |v_i|^{\alpha - 2}$$

Обработка многократно измеренных величин

Альтернативные оценки сдвига (положения, МО):

- самая старая (*Менделеев, Пуанкаре*) - *усеченное среднее* (α-усеченное среднее), со степенью усечения α%. При вычислении отбрасывают $[(\alpha\%)\cdot n]$ мин и мах элементов в ряде из оставшихся среднее; Поздний аналог - *винзоризованное среднее* (α- винзоризованное *среднее Винзора*).

Усеченное средне - потеря информации, *Винзоризованное среднее* - если экстремальные элементы грубые они тиражируются.

Степень усечения не известна - обычно принимают равным 10%. Осторожное использование.

Обработка многократно измеренных величин

Альтернативные оценки сдвига (положения, МО):

- самая известная и распространенная (почему?) *медиана* ряда ($med(x)$, $Me(x)$, L_1-*оценка*). Это средний элемент в вариационном ряду.

Возможно через коэффициент

$$k = (n + 1)/2.$$

Медиана – наиболее устойчивая оценка к наличию грубых измерений и достаточно устойчива к некоторым видам систематического влияния (*Эджворт*). Не учитывает (отсекает) значения элементов.

Основа многих робастных оценок.

Обработка многократно измеренных величин

- самая известная оценка на основе *медианы* - *псевдомедиана*, (*R-оценка Ходжеса-Лемана*). Это *медиана* из всех возможных пар средних i и j вперед ($i \geq j$) из вариационного ряда (не обязательно).

Попарные средние в статистике - средние *Уолша*

$$(\overline{x}_W)_{ij} = \frac{x_i + x_j}{2}$$

Псевдомедиана (оценка *Ходжеса-Лемана*) ряда будет

$$\theta_{HL} = med(\overline{x}_W) = med\left(\frac{(x_i + x_j)}{2}\right),\ i \geq j$$

Модификации $i > j$, медиана подразмахов Бикел-Ходжеса

Обработка многократно измеренных величин

- *адаптивная оценка Хогга*. основа - вычисление индикатора по исследованию ряда и ветвление по схеме

$$\theta_{AH} = \begin{cases} S(0.25;n), & k < 2; \\ \tilde{N}_t(0;n), & 2 < k \le 4; \\ \tilde{N}_t(0.25;n), & 4 < k \le 5.5; \\ \tilde{N}_t(0.5;n), & 5.5 < k. \end{cases}$$

S и *C* – *среднее* из *t*% первых или последних значений вариационного ряда;

Очень устойчива к отклонениям (нормальность, случайность, однородность и независимость), достаточно эффективна уже при числе $n > 5$, одна из лучших альтернативных оценок.

Обработка многократно измеренных величин

Чем пользоваться - коэффициент k (индикатор). Характеризует «тяжесть хвостов» распределения.

Для оценки коэффициента k используют

– значение оценки не приведенного *эксцесса*

$$k = E^{*} = \frac{[(x - \overline{x})^{4}]}{n \cdot \sigma^{4}}$$

Для точно нормального закона равна 3; Неустойчив.

– устойчивая оценка коэффициента, на основе линейных комбинаций измерений (*Хогг*)

$$k = t_{n} = \frac{(a_{n}(0.05) - b_{\mathrm{n}}(0.05))}{(a_{n}(0.5) - b_{\mathrm{n}}(0.5))}$$

Обработка многократно измеренных величин

Здесь $a_n(\beta)$, $b_n(\beta)$ - *средние* по $(100\cdot\beta)\%$ наибольших и наименьших элементов *вариационного ряда* соответственно

Общая схема для эффективного оценивания *сдвига* ряда измерений:

- *среднее арифметическое*

- вычисляется одна (или несколько) альтернативных робастных оценок.

- если нет отличия (минимально), оставляют среднее арифметическое. Отличия большие - окончательная оценка только альтернативная.

Обработка многократно измеренных величин

Альтернативные оценки масштаба:

- известна ещё *Гауссу*, *средняя абсолютная погрешность* (*САП*(*x*))

$$САП(x) = \frac{\left[\left|x - \overline{x}\right|\right]}{n}$$

Отдавал предпочтение *Фишер* по сравнению с *стандартным отклонением* σ (дискуссия Фишер-Эддингтон). Теоретически отношение *стандартного отклонения* к *средней абсолютной погрешности*

$$k_1 = \frac{\sigma}{САП(x)} = \sqrt{\frac{\pi}{2}} = 1.2533... \quad \text{и} \quad \sigma = 1.2533 \cdot САП(x)...$$

Обработка многократно измеренных величин

- известна ещё *Гауссу абсолютное медианное отклонение* (АМО, *вероятная ошибка*) – медиана абсолютных отклонений от медианы

$$AMO(x) = med(|x - med(x)|)$$

Все положительные свойства медианы (и отрицательные). Невероятно робастна - основа многих других оценок масштаба. Теоретическое отношение *стандартного отклонения* к *абсолютному медианному отклонению*

$$k_2 = \frac{\sigma}{AMO(x)} = \frac{1}{F^{-1}(0.75)} = 1.4826...$$

и

$$\sigma = 1.4826 \cdot AMO(x)$$

Обработка многократно измеренных величин

- *интерквартильный размах* (*IQR*(*x*)) - R-оценка, эффективная и устойчивая оценка. Основа - *квартили* (*четверти*).

p-квартиль - граничное число вариационного ряда, относительно которого вероятность появления значений меньше этой границы равна *p*. Четверти - *p* есть 25%, 50% и 75%.

Первая четверть Q_1 (*нижняя*, или *первая квартиль*), вторая Q_2 (средняя, или *вторая квартиль*, *медиана*), третья четверть Q_3 (*верхняя*, или *третья квартиль*).

Много способов вычисления *квартилей* Q_i

и

Обработка многократно измеренных величин

Одна из неплохих схем:

– строят вариационный ряд и считают коэффициент

$$t = \frac{n+1}{100}$$

– номер элемента в вариационном ряде из n элементов, соответствующий нужной квантили вычисляют как

$$n_{25} = \left(\frac{n+1}{100}\right) \cdot 25 = \frac{n+1}{4},$$

$$n_{50} = \left(\frac{n+1}{100}\right) \cdot 50 = \frac{n+1}{2},$$

$$n_{75} = \left(\frac{n+1}{100}\right) \cdot 75 = \frac{3}{4} \cdot (n+1)$$

Обработка многократно измеренных величин

Часто номер дробный ($C.D$)- линейная интерполяция:

$$Q_i = X_{(C)} + D \cdot \left(X_{(C+1)} - X_{(C)} \right) = X_{(C)} \cdot (1 - D) + D \cdot X_{(C+1)}$$

Интерквартильный размах

$$IQR = Q_3 - Q_1$$

Теоретическое отношение *стандартного отклонения* к *интерквартильному размаху*

$$k_3 = \frac{\sigma}{IQR(x)} = \frac{1}{2F^{-1}(0.75)} = 0.7413...$$

и

$$\sigma = 0.7413 \cdot IQR(x)$$

Обработка многократно измеренных величин

– *оценка Даунтона* (1966 г.), *L*-оценка, надёжна, устойчива к нарушению основных предположений и очень эффективна

$$\theta_D = \frac{2\sqrt{\pi}}{n(n-1)} \sum_{i=1}^{n} \left(\left(i - \frac{n+1}{2} \right) \cdot x_{(i)} \right) =$$

$$= \frac{1.7724}{n(n-1)} \sum_{i=1}^{n} \left((2i - n - 1) \cdot x_{(i)} \right)$$

или

$$\theta_D = \frac{1.7724}{n(n-1)} \sum_{\substack{i,j=1 \\ j>i}}^{n} \left| d_{ij} \right| =$$

$$= \frac{1.7724}{n(n-1)} \sum_{i=1}^{\left[\frac{n}{2}\right]} \left((n - 2i + 1) \cdot w_i \right)$$

через последовательные разности или подразмахи.

Обработка многократно измеренных величин

Параметризации:

- Усреднение
- Медианирование
- Взвешивание.

Статистики:

- Натуральные
- Порядковые
- Средние Уолша
- Разности Джини
- Поразмахи

Обработка многократно измеренных величин

Оценка по невязкам. Невязки – истинные погрешности:

$$\mu = \sqrt{\frac{[pw^2]}{n}}$$

Формула Ферреро.

$$w_i = (\beta_1 + \beta_2 + \beta_3) - 180$$

$$p_{w_i} = \frac{1}{3}$$

$$\frac{1}{p_w} = 1^2 \cdot \frac{1}{p_1} + 1^2 \cdot \frac{1}{p_2} + 1^2 \cdot \frac{1}{p_3}$$

Варианты. СВ.

$$\mu = \sqrt{\frac{[pw^2]}{n}} = \sqrt{\frac{\frac{1}{3}[w^2]}{n}} = \sqrt{\frac{[w^2]}{3n}} = m_\beta$$

Обработка многократно измеренных величин

Вопросы по теории модуля 2:

1. Оценивание одной многократно измеренной величины. Общие положения.
2. Оценивание одной многократно измеренной равноточной величины классически.
3. Оценивание одной многократно измеренной неравноточной величины классически.
4. Исследование отклонений от основных предположений.
5. Альтернативные оценки сдвига
6. Альтернативные оценки масштаба