Меры вариабельности оценок параметров в задачах восстановления зависимостей по интервальным данным

С.П. Шарый

ФИЦ ИВТ и НГУ, Новосибирск, Россия

MMOTI-2021, 16 ноября 2021 г.

Введение и постановка задачи

Решаем задачу восстановления зависимостей:

> для набора результатов измерений или наблюдений необходимо построить функцию заданного вида, которая «лучше всего подходит» этим данным.

Введение и постановка задачи

Решаем задачу восстановления зависимостей:

для набора результатов измерений или наблюдений необходимо построить функцию заданного вида, которая «лучше всего подходит» этим данным.

Определить параметры $\beta_1, \beta_2, \ldots, \beta_n$ линейной функции

$$y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

из множества значений независимых переменных $x_1, x_2, \ldots, x_n$ и соответствующих значений зависимой переменной y.

Как x_1, x_2, ..., x_n, так и y не известны точно, и ,

в i-ом измерении мы имеем только интервалы

$$x_1 \in [\underline{x}_{i1}, \overline{x}_{i1}] = \boldsymbol{x}_{i1},\ x_2 \in [\underline{x}_{i2}, \overline{x}_{i2}] = \boldsymbol{x}_{i2},\ \dots,\ x_n \in [\underline{x}_{in}, \overline{x}_{in}] = \boldsymbol{x}_{im}$$

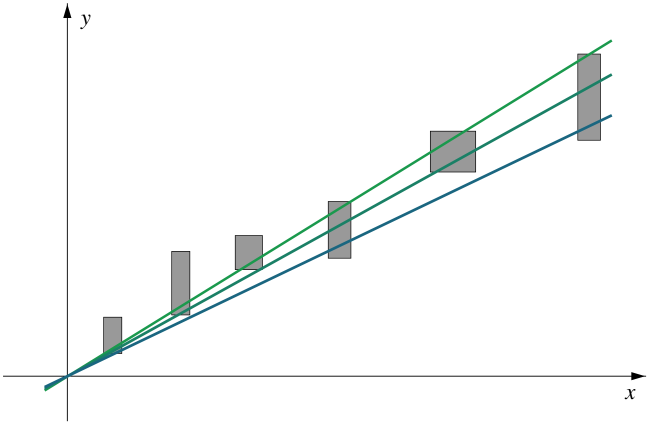
$$y \in \left[\, \underline{y}_i\,, \overline{y}_i \,\right] = \boldsymbol{y}_i, \qquad i = 1, 2, \dots, m.$$

Исходными данными задачи является набор интервалов значений

независимых и зависимых переменных нашей функции:

$$\begin{matrix} \boldsymbol{x}_{11}, & \boldsymbol{x}_{12}, & \dots & \boldsymbol{x}_{1n}, & \boldsymbol{y}_1, \\ \boldsymbol{x}_{21}, & \boldsymbol{x}_{22}, & \dots & \boldsymbol{x}_{2n}, & \boldsymbol{y}_2, \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{x}_{m1}, & \boldsymbol{x}_{m2}, & \dots & \boldsymbol{x}_{mn}, & \boldsymbol{y}_m. \end{matrix}$$

Иллюстрация



Для нахождения оценок коэффициентов β_1, β_2, ..., β_n,

разработаны несколько методов.

Метод максимума совместности (“метод максимума согласования”):

- Шарый С.П. Разрешимость интервальных линейных уравнений и анализ данных с неопределённостью, *Автоматика и Телемеханика*, №2 (2012), сс. 111–125.
- Shary, S.P. Maximum consistency method for data fitting under interval uncertainty, *Journal of Global Optimization*, vol. 66 (2016), pp. 111–126.
- Shary, S.P. Weak and strong compatibility in data fitting under interval uncertainty, *Advances in Data Science and Adaptive Analysis*, vol. 12 (2020), No. 1, 2050002.

Введение и постановка задачи

Термином «вариабельность»

мы называем степень изменчивости и неоднозначности оценки.

Обычно мы получаем целое множество различных оценок, которые одинаково согласуются (совместны) с исходными данными и поэтому подходят в качестве решений задачи.

Количественная характеристика того, насколько это множество

велико или мало, будет называться «вариабельностью».

Мера вариабельности

В традиционной вероятностной статистике

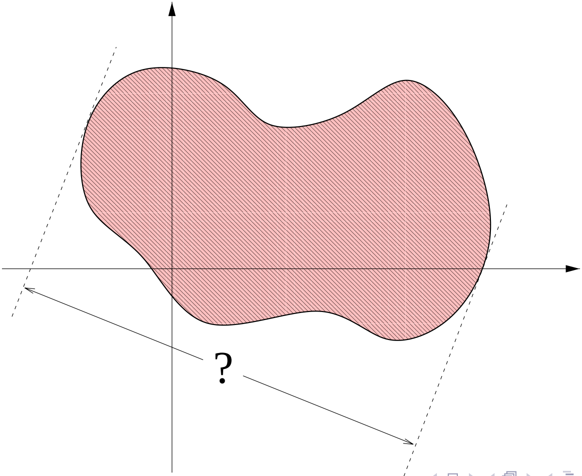
- стандартное отклонение,
- среднее абсолютное отклонение,
- коэффициент Джини,
- медианное абсолютное отклонение,

и другие подобные меры используются для описания вариабельности и рассеяния оценок параметров.

Что может быть их аналогом в Интервальном Анализе Данных?

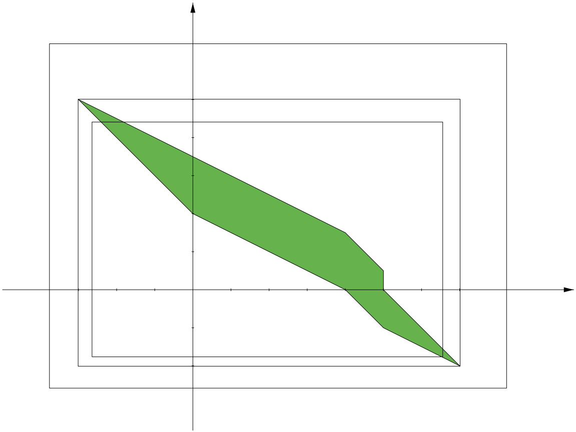
Мера вариабельности

Мера вариабельности может быть оценкой размеров множества решений задачи



Мера вариабельности

Мы можем даже взять внешнюю оценку множества решений, полученную каким-либо интервальным методом.



Мера вариабельности

Мы можем даже взять внешнюю оценку множества решений, полученную каким-либо интервальным методом.

Недостатки:

- излишняя детализация ответа, выдаваемого как интервальный брус в $\mathbb{R}^n$, большой объём информации, который ещё нужно «переварить» и редуцировать к компактной форме.
- вычислительная сложность получения такой оценки.

Мера вариабельности

Нужно иметь относительно простую и эффективно вычислимую оценку, выражаемую одним числом, которая давала бы общий, «аггрегированный», взгляд.

Аналогично дисперсии и другим вероятностным мерам, она могла бы служить приближённой характеристикой качества оценки параметров.

Чем больше вариабельность оценки,

тем хуже её определённость и однозначность.

Метод максимума совместности

Исходными данными в нашей задаче является набор значений независимых и зависимых переменных функции:

$$
\begin{array}{ccccc}
\boldsymbol{x}_{11}, & \boldsymbol{x}_{12}, & \dots & \boldsymbol{x}_{1n}, & \boldsymbol{y}_1, \\
\boldsymbol{x}_{21}, & \boldsymbol{x}_{22}, & \dots & \boldsymbol{x}_{2n}, & \boldsymbol{y}_2, \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\boldsymbol{x}_{m1}, & \boldsymbol{x}_{m2}, & \dots & \boldsymbol{x}_{mn}, & \boldsymbol{y}_m.
\end{array}
$$

Метод максимума совместности

Исходными данными в нашей задаче является набор значений независимых и зависимых переменных функции:

$$\begin{matrix} \boldsymbol{x}_{11}, & \boldsymbol{x}_{12}, & \ldots & \boldsymbol{x}_{1n}, & \boldsymbol{y}_1, \\ \boldsymbol{x}_{21}, & \boldsymbol{x}_{22}, & \ldots & \boldsymbol{x}_{2n}, & \boldsymbol{y}_2, \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{x}_{m1}, & \boldsymbol{x}_{m2}, & \ldots & \boldsymbol{x}_{mn}, & \boldsymbol{y}_m. \end{matrix}$$

Конструируемая линейная функция

$$y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

Чтобы найти оценку параметров β_1, β_2, ..., β_n, мы «подставляем» данные в равенство для функции, получая интервальную систему уравнений

$$
\left\{
\begin{array}{ccccccccc}
\boldsymbol{x}_{11}\beta_1 & + & \boldsymbol{x}_{12}\beta_2 & + & \ldots & + & \boldsymbol{x}_{1n}\beta_n & = & \boldsymbol{y}_1, \\
\boldsymbol{x}_{21}\beta_1 & + & \boldsymbol{x}_{22}\beta_2 & + & \ldots & + & \boldsymbol{x}_{2n}\beta_n & = & \boldsymbol{y}_2, \\
\vdots & & \vdots & & \ddots & & \vdots & & \vdots \\
\boldsymbol{x}_{m1}\beta_1 & + & \boldsymbol{x}_{m2}\beta_2 & + & \ldots & + & \boldsymbol{x}_{mn}\beta_n & = & \boldsymbol{y}_m,
\end{array}
\right.
$$

или, кратко,

$$
\boldsymbol{X}\beta = \boldsymbol{y}
$$

с интервальной $m \times n$-матрицей $\boldsymbol{X} = (\boldsymbol{x}_{ij})$

и интервальным m-вектором $\boldsymbol{y} = (\boldsymbol{y}_i)$.

Параметры, которые совместны, в том или ином смысле, с данными измерений, образуют множество решений для интервальной системы уравнений.

Параметры, которые совместны, в том или ином смысле, с данными измерений, образуют множество решений для интервальной системы уравнений.

Объединённое множество решений,

$$\Xi_{uni}(\boldsymbol{X}, \boldsymbol{y}) = \{\, \beta \in \mathbb{R}^n \mid (\exists X \in \boldsymbol{X})(\exists y \in \boldsymbol{y})(X\beta = y) \,\},$$

соответствует «слабой совместности» между параметрами и данными.

Допусковое множество решений,

$$\Xi_{tol}(\boldsymbol{X}, \boldsymbol{y}) = \{\, \beta \in \mathbb{R}^n \mid (\forall X \in \boldsymbol{X})(\exists y \in \boldsymbol{y})(X\beta = y) \,\},$$

отвечает «сильной совместности» между параметрами и данными.

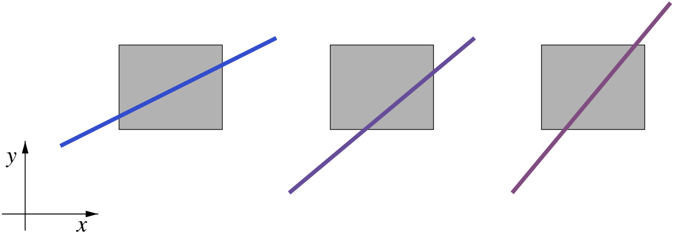
Метод максимума совместности

Вводим специальную «меру совместности» вектора параметров
с эмпирическими данными — *распознающий функционал*.

В качестве оценки параметров берём точку, в которой
достигается максимум распознающего функционала.

Метод максимума совместности
имеет две версии,
«слабую» и «сильную».

Слабая и сильная совместность отражают две различные ситуации которые могут встретиться при обработке данных:



Слабая и сильная совместность отражают две различные ситуации которые могут встретиться при обработке данных:

- В слабой версии требуется, чтобы график конструируемой функции просто пересекал брусы неопределённости данных.

- В сильной версии требуется, чтобы график функции проходил через «коридоры», задаваемые интервалами $\boldsymbol{y}_i$, $i = 1, 2, \ldots, m$, для *любого* значения независимых переменых x_1, x_2, ..., x_n из интервалов $\boldsymbol{x}_{i1}$, $\boldsymbol{x}_{i2}$, ..., $\boldsymbol{x}_{in}$, полученных в i-ом измерении.

Иллюстрация «слабой совместности»

между интервальными данными и линейной функцией

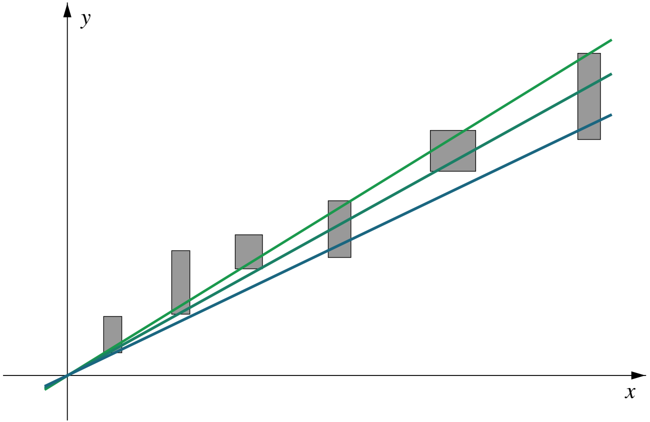
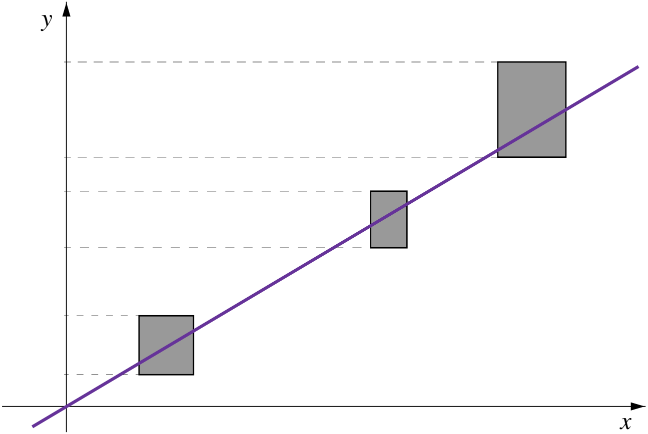


Иллюстрация «сильной совместности» между интервальными данными и линейной функцией



Метод максимума совместности

«Сильная версия» метода максимума совместности
имеет ряд преимуществ перед «слабой версией»:

- полиномиальная сложность,
- робастность оценок,
- конечная вариабельность оценок,
- сильная совместность частично преодолевает так называемый парадокс Демиденко, и т. .д.

Рассматриваем сильную версию метода максимума совместности, которая соответствует допусковому множеству решений $\Xi_{tol}(\boldsymbol{X}, \boldsymbol{y})$.

Полезный результат о допусковом множестве решений —

Критерий ограниченности И.А. Шарой

Пусть для интервальной линейной системы $\boldsymbol{X}\beta = \boldsymbol{y}$ допусковое множество решений непусто. Оно неограниченно тогда и только тогда, когда матрица $\boldsymbol{X}$ имеет линейно зависимые неинтервальные столбцы.

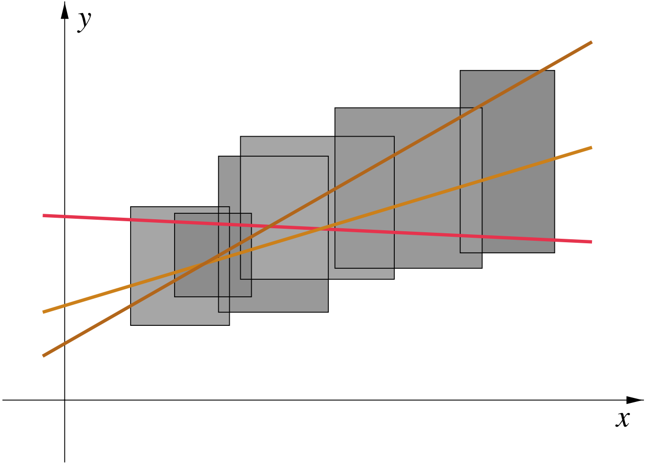
Полезный результат о допусковом множестве решений —

Критерий ограниченности И.А. Шарой

Пусть для интервальной линейной системы $\boldsymbol{X}\beta = \boldsymbol{y}$ допусковое множество решений непусто. Оно неограниченно тогда и только тогда, когда матрица $\boldsymbol{X}$ имеет линейно зависимые неинтервальные столбцы.

Допусковое множество решений ИСЛАУ почти всегда ограниченно.

Поэтому оценки, полученные сильной версией метода максимума совместности, почти всегда имеют конечную вариабельность.


y
x

Распознающий функционал допускового множества решений,
который даёт меру сильной совместности:

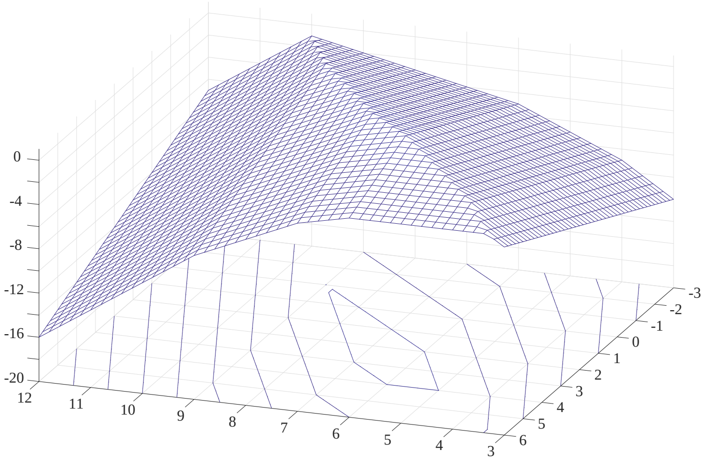
$$\mathrm{Tol}\,(\beta, \boldsymbol{X}, \boldsymbol{y}) \;=\; \min_{1\le i\le m} \left\{ \mathrm{rad}\,\boldsymbol{y}_i - \left| \,\mathrm{mid}\,\boldsymbol{y}_i - \sum_{j=1}^{n} \boldsymbol{x}_{ij}\beta_j \right| \right\},$$

где

$$\mathrm{rad}\,\boldsymbol{y}_i = \tfrac{1}{2}(\overline{\boldsymbol{y}}_i - \underline{\boldsymbol{y}}_i), \qquad \mathrm{mid}\,\boldsymbol{y}_i = \tfrac{1}{2}(\overline{\boldsymbol{y}}_i + \underline{\boldsymbol{y}}_i),$$

$$|\boldsymbol{a}| = \max\{\,|a| \mid a \in \boldsymbol{a}\,\} = \max\{\,|\underline{\boldsymbol{a}}|, |\overline{\boldsymbol{a}}|\,\}.$$

Типичный график распознающего функционала Tol



Метод максимума совместности

Чтобы решить задачу восстановления линейной зависимости, необходимо найти безусловный максимум, по всем $\beta \in \mathbb{R}^n$, функционала Tol,

$$\mathrm{Tol}\,(\beta, \boldsymbol{X}, \boldsymbol{y}) \to \max,$$

и вектор

$$\hat{\beta} = \arg\max_{\beta \in \mathbb{R}^n} \mathrm{Tol}\,(\beta, \boldsymbol{X}, \boldsymbol{y})$$

даёт оценку параметров линейной функции.

Тогда $\hat{\beta}$ называется *оценкой максимума совместности*.

Метод максимума совместности и допусковое множество решений

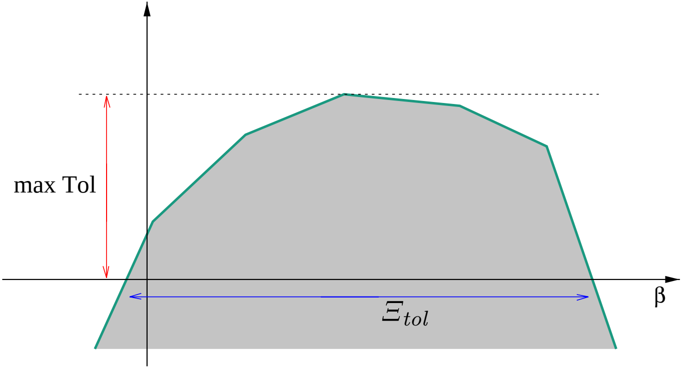
- Если $\max \operatorname{Tol} \geq 0$, то множество решений $\Xi_{tol}(\boldsymbol{X}, \boldsymbol{y})$, т. е. множество параметров, сильно совместных с данными, непусто, и $\hat{\beta} \in \Xi_{tol}(\boldsymbol{X}, \boldsymbol{y})$.

- Если $\max \operatorname{Tol} < 0$, то множество решений $\Xi_{tol}(\boldsymbol{X}, \boldsymbol{y})$ пусто и не существует параметров, сильно совместных с данными.

  Тем не менее, аргумент $\hat{\beta}$ of $\max \operatorname{Tol}$ всё-таки обеспечивает наилучшую возможную совместность построенной функции с данными (наименьшую несовместность).

Мера вариабельности

Идея



Максимальное значение распознающего функционала
даёт информацию о размерах допускового множества решений Ξ_{tol}.

Мера вариабельности

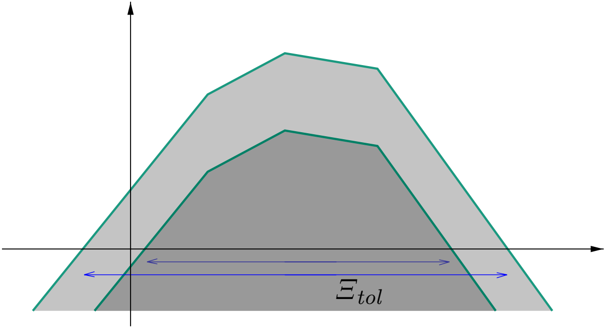
$$\Xi_{tol}(\boldsymbol{X}, \boldsymbol{y}) = \{\, \beta \in \mathbb{R}^n \mid \mathrm{Tol}\,(\beta, \boldsymbol{X}, \boldsymbol{y}) \geq 0 \,\}$$

Величина $\max\, \mathrm{Tol}$ может, при прочих равных условиях, быть взята мерой того, насколько мало или велико множество решений Ξ_{tol}.

Чем больше $\max\, \mathrm{Tol}$,

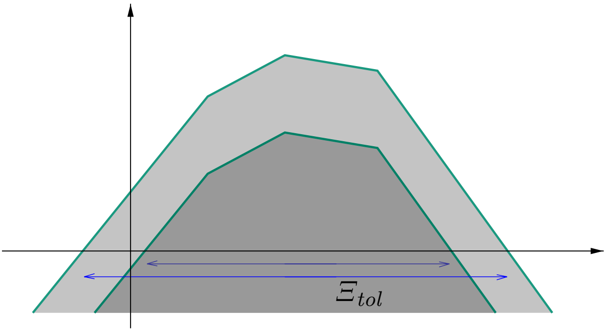
тем больше размеры допускового множества решений,

и наоборот.



Размеры множества решений пропорциональны $\max \operatorname{Tol}$

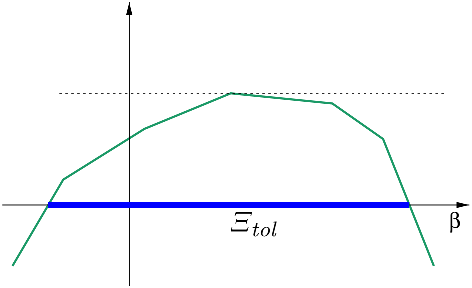
для достаточно больших $\operatorname{rad} \boldsymbol{y}$,

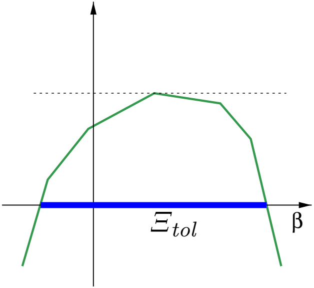


Размеры множества решений пропорциональны $\max \mathrm{Tol}$

для достаточно больших $\mathrm{rad}\, \boldsymbol{y}$,

$$\boxed{\text{мера вариабельности} \; = \; K \cdot \max \mathrm{Tol}}$$





В дополнение к $\max \mathrm{Tol}$

размеры Ξ_{tol} также определяются

крутизной (наклоном) графика Tol.

Мера вариабельности

В качестве характеристики вариабельности оценки параметров для линейной функции, $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_n)$, получаемой с помощью метода максимума совместности, мы предлагаем

$$\mathrm{IVE}\,(\boldsymbol{X}, \boldsymbol{y}) \;=\; \sqrt{n}\;\; \max_{\mathbb{R}^n} \,\mathrm{Tol} \cdot \Big(\min_{X \in \mathrm{vert}\,\boldsymbol{X}} \,\mathrm{cond}_2\, X \Big) \cdot \frac{\big\| \arg\max\limits_{\mathbb{R}^n} \,\mathrm{Tol} \,\big\|_2}{\|\hat{\boldsymbol{y}}\|_2}$$

В выписанной формуле

n — размерность вектора параметров функции;

$\|\cdot\|_2$ — евклидова норма (2-норма) векторов из $\mathbb{R}^n$,

$$\|a\|_2 \;=\; \Bigl(\,\textstyle\sum_{i=1}^{n} |a_i|^2\,\Bigr)^{1/2};$$

$\operatorname{vert}\boldsymbol{X}$ — множество «концевых матриц» для интервальной матрицы $\boldsymbol{X}$, т. е. множество таких точечных матриц $X = (x_{ij})$, что $x_{ij} \in \{\underline{\boldsymbol{x}}_{ij}, \overline{\boldsymbol{x}}_{ij}\}$;

$\mathrm{cond}_2 X$ — спектральное число обусловленности матрицы,

$$\mathrm{cond}_2\, X \;=\; \sigma_{\max}(X)\,/\,\sigma_{\min}(X);$$

$\hat{\boldsymbol{y}}$ — «наиболее представительная» точка из интервального вектора $\boldsymbol{y}$, которая берётся как

$$\hat{\boldsymbol{y}} \;=\; \tfrac{1}{2}(|\mathrm{mid}\,\boldsymbol{y} + \mathrm{rad}\,\boldsymbol{y}| + |\mathrm{mid}\,\boldsymbol{y} - \mathrm{rad}\,\boldsymbol{y}|).$$

Мера вариабельности

Наклон гиперплоскостей определяется коэффициентами уравнений, которые задают их, и они являются концами интервалов данных.

Этот наклон обобщённо выражается числом обусловленности точечных матриц из интервальной матрицы данных $\boldsymbol{X}$.

Наконец, множитель

$$\frac{\|\arg\max \mathrm{Tol}\|_2}{\|\hat{\boldsymbol{y}}\|_2} = \frac{\|\hat{\beta}\|_2}{\|\hat{\boldsymbol{y}}\|_2}$$

— это масштабирующий коэффициент, котрый обеспечивает соизмеримость окончательного значения с величиной решения и вектором правой части системы.

Таким образом, получаем формулу для IVE.

Обоснование

Sergey P. Shary, A variability measure for estimates of parameters in interval data fitting, arXiv:2003.05126

Возмущение Δx ненулевого решения x

для системы линейных уравнений $Ax = b$:

$$\frac{\|\Delta x\|_2}{\|x\|_2} \leq \operatorname{cond}_2 A \cdot \frac{\|\Delta b\|_2}{\|b\|_2}.$$

Это неравенство справедливо также для переопределённых систем.

Обоснование

Sergey P. Shary, A variability measure for estimates of parameters in interval data fitting, arXiv:2003.05126

Возмущение Δx ненулевого решения x

для системы линейных уравнений $Ax = b$:

$$\frac{\|\Delta x\|_2}{\|x\|_2} \;\leq\; \operatorname{cond}_2 A \cdot \frac{\|\Delta b\|_2}{\|b\|_2}.$$

Это неравенство справедливо также для переопределённых систем.

Тогда

$$\|\Delta x\|_2 \;\leq\; \|\Delta b\|_2 \cdot \operatorname{cond}_2 A \cdot \frac{\|x\|_2}{\|b\|_2},$$

где $\|\Delta b\|_2 \;=\; \sqrt{n}\;\max \mathrm{Tol}\,.$

Вычисление IVE

$$\mathrm{IVE}\,(\boldsymbol{X},\boldsymbol{y}) \;=\; \sqrt{n}\;\;\max_{\mathbb{R}^n}\,\mathrm{Tol}\cdot\left(\min_{X\in\mathrm{vert}\,\boldsymbol{X}}\;\mathrm{cond}_2\,X\right)\cdot\frac{\left\|\,\arg\max\limits_{\mathbb{R}^n}\,\mathrm{Tol}\,\right\|_2}{\|\hat{\boldsymbol{y}}\|_2}$$

основная трудность — в вычислении $\min_{X\in\mathrm{vert}\,\boldsymbol{X}}\;\mathrm{cond}_2\,X$

Вычисление IVE

$$\mathrm{IVE}\,(\boldsymbol{X},\boldsymbol{y}) \;=\; \sqrt{n}\;\max_{\mathbb{R}^n}\,\mathrm{Tol}\cdot\left(\min_{X\in\mathrm{vert}\,\boldsymbol{X}}\;\mathrm{cond}_2\,X\right)\cdot\frac{\left\|\arg\max\limits_{\mathbb{R}^n}\,\mathrm{Tol}\right\|_2}{\|\hat{\boldsymbol{y}}\|_2}$$

основная трудность — в вычислении $\min_{X\in\mathrm{vert}\,\boldsymbol{X}}\;\mathrm{cond}_2\,X$

Если $\boldsymbol{X}$ — достаточно узкая, то

$$\min_{X\in\mathrm{vert}\,\boldsymbol{X}}\;\mathrm{cond}_2\,X \;\approx\; \mathrm{cond}_2(\mathrm{mid}\,\boldsymbol{X}).$$

В общем случае можем использовать эволюционные алгоритмы оптимизации, такие как генетический алгоритм, симулированный отжиг, метод «роя частиц» и т. п.

Мера вариабельности

В качестве характеристики вариабельности оценки параметров линейной функции, $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_n)$, получаемой с помощью метода максимума совместности, мы предлагаем

$$\mathrm{IDE}\,(\boldsymbol{X}, \boldsymbol{y}) \;=\; \sqrt{n}\;\; \max_{\mathbb{R}^n} \mathrm{Tol} \cdot \mathrm{cond}_2\, \mathcal{X} \cdot \frac{\bigl\| \arg\max\limits_{\mathbb{R}^n} \mathrm{Tol} \bigr\|_2}{\|\hat{\boldsymbol{y}}\|_2}.$$

В выписанной формуле

n — размерность вектора параметров функции;

$\|\cdot\|_2$ — евклидова норма (2-норма) векторов в $\mathbb{R}^n$;

$\mathcal{X}$ — специальная матрица, называемая *матрицей поконцевых комбинаций* для $\boldsymbol{X}$, имеющая размеры $N \times n$ с $N \leq m \cdot 2^n$, составленная из комбинаций концов интервальных элементов вдоль каждой строки $\boldsymbol{X}$;

$\mathrm{cond}_2 X$ — спектральное число обусловленности матрицы;

$\hat{\boldsymbol{y}}$ — «наиболее представительная» точка из $\boldsymbol{y}$,

$$\hat{\boldsymbol{y}} \;=\; \tfrac{1}{2}(|\mathrm{mid}\,\boldsymbol{y} + \mathrm{rad}\,\boldsymbol{y}| + |\mathrm{mid}\,\boldsymbol{y} - \mathrm{rad}\,\boldsymbol{y}|).$$

Вычисление IDE

$$\mathrm{IDE}\,(\boldsymbol{X},\boldsymbol{y}) \;=\; \sqrt{n}\;\max_{\mathbb{R}^n}\,\mathrm{Tol}\cdot\mathrm{cond}_2\,\mathcal{X}\cdot\frac{\left\|\,\arg\max\limits_{\mathbb{R}^n}\;\mathrm{Tol}\,\right\|_2}{\|\hat{\boldsymbol{y}}\|_2}.$$

— основная трудность заключается в конструировании матрицы поконцевых комбинаций $\mathcal{X}$ с $m\cdot 2^n$ строками.

Вычисление IDE

$$\mathrm{IDE}\,(\boldsymbol{X},\boldsymbol{y}) \;=\; \sqrt{n}\;\;\max_{\mathbb{R}^n}\,\mathrm{Tol}\,\cdot\,\mathrm{cond}_2\,\mathcal{X}\,\cdot\,\frac{\left\|\,\arg\max\limits_{\mathbb{R}^n}\,\mathrm{Tol}\,\right\|_2}{\|\hat{\boldsymbol{y}}\|_2}.$$

— основная трудность заключается в конструировании матрицы поконцевых комбинаций $\mathcal{X}$ с $m\cdot 2^n$ строками.

Естественный выход из затруднения — использовать «редуцированную матрицу поконцевых комбинаций» $\tilde{\mathcal{X}}$.

$\tilde{\mathcal{X}}$ конструируется специальным образом и обычно имеет $O(2m)$ строк, составленных из некоторых, не всех, комбинаций концов по строкам $\boldsymbol{X}$

Редуцированная матрица поконцевых комбинаций

$$\boxed{\beta \in \Xi_{tol}(\boldsymbol{X}, \boldsymbol{y}) \qquad \Longleftrightarrow \qquad \boldsymbol{X} \cdot \beta \subseteq \boldsymbol{y}}$$

Если мы знаем знаки компонент $\beta = (\beta_1, \beta_2, \ldots, \beta_n)^\top$, то

$$(\boldsymbol{X} \cdot \beta)_i \;=\; \left[\; \sum_{i=1}^{n} x'_{ij}\beta_j, \sum_{i=1}^{n} x''_{ij}\beta_j \;\right]$$

с некоторыми определёнными наборами концов $x'_{ij}, x''_{ij} \in \{\, \underline{\boldsymbol{x}}_{ij}, \overline{\boldsymbol{x}}_{ij} \,\}$.

Тогда можем уменьшить число неравенств

в системе, равносильной $\boldsymbol{X} \cdot \beta \subseteq \boldsymbol{y}$.

Спасибо за внимание