

ИНТЕРВАЛЬНЫЙ ПОДХОД К ОБРАБОТКЕ ЗАШУМЛЕННЫХ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ С МНОГОКРАТНЫМИ ИЗМЕРЕНИЯМИ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

С. И. Кумков

Отдел динамических систем

Институт математики и механики УрО РАН, Екатеринбург

kumkov@imm.uran.ru

Международная Конференция “Современные проблемы прикладной математики и механики: теория, эксперимент и практика”, посвящённая 90-летию со дня рождения академика Н.Н.Яненко

ИВТ СО РАН, Новосибирск, Россия, 30 мая – 4 июня 2011 г.

Литература

Задачи прикладной химии и исходные данные

1. Лукьянова М.Я., Осипенко А.Г., Маершин А.А., Кормилицин М.В., Кумков С.И., Валишин М.Ф. Задача обработки экспериментальных данных по хронопотенциометрии кюрия в расплаве $3\text{LiCl}-2\text{KCl}$ // Доклады III Международной Пирохимической Конференции, 29.11–03.12. 2010, Димитровград, Россия, Исследовательский институт атомных реакторов.

2. Вакарин С.В., Микушина Ю.В., и др. Каталазная активность поликристаллических и наноразмерных оксидных вольфрамовых бронз, полученных электролизом расплавленных солей // Институт органического синтеза УрО РАН, Институт высокотемпературной электрохимии УрО РАН, Екатеринбург, 2011.

3. Новоселова А.В., Смоленский В.В. Окислительно-восстановительные потенциалы $\text{Tm(III)}/\text{Tm(II)}$ в расплавленном хлориде цезия // Труды Новомосковского института Российского химико-технологического университета им. Д.И. Менделеева. Серия Физическая химия и электрохимия / Под ред. проф. А.В. Волковича. Новомосковск: ГОУ ВПО НИ РХТУ. 2009. вып. 3 (23). С. 142-154.

Стандартные статистические методы обработки экспериментальных данных

4. ГОСТ 8.207-76. Государственная система обеспечения единства измерений. Прямые измерения с многократными наблюдениями. Методы обработки результатов наблюдений.

5. МИ 2083-93. Рекомендации. Государственная система обеспечения единства измерений. Измерения косвенные. Определение результатов измерений и оценивание их погрешностей.

Литература

6. Р 40.2.028–2003. Государственная система обеспечения единства измерений. Рекомендации по построению градуировочных характеристик. Оценивание погрешностей (неопределённости) линейных градуировочных характеристик при использовании метода наименьших квадратов.

Основы и прикладные вопросы интервального анализа

7. Шарый С.П. Конечномерный интервальный анализ. Электронная книга:
<http://www.nsc.ru/interval/Library/InteBooks/SharyBook.pdf>

8. Шарый С.П. Разрешимость интервальных линейных уравнений и анализ данных с неопределённостями // Сб. Тезисов XIV-й Всероссийской конференции “Математическое программирование и приложения”, г. Екатеринбург, 28 февраля - 4 марта 2011 года. ИММ УрО РАН, г. Екатеринбург, 2011.

9. Жолен Л., Кифер М., Дидри О., Вальтер Э. Прикладной интервальный анализ. Москва–Ижевск, Регулярная и хаотическая динамика, 2007.

10. Фидлер М., Недома Й., Рамик Я., Рон И., Циммерманн К. Задачи линейной оптимизации с неточными данными. Москва–Ижевск, Регулярная и хаотическая динамика, 2008.

11. Хансен Э., Дж.У. Уолстер. Глобальная оптимизация с помощью методов интервального анализа. Москва–Ижевск, Регулярная и хаотическая динамика, 2010.

Литература

12. Кумков С.И. Решение систем неравенств в условиях неопределённости. Методы интервального анализа // Сб. Тезисов XIV-й Всероссийской конференции “Математическое программирование и приложения”, г. Екатеринбург, 28 февраля - 4 марта 2011 года. ИММ УрО РАН, г. Екатеринбург, 2011.

13. Кумков С.И. Обработка экспериментальных данных ионной проводимости расплавленного электролита методами интервального анализа // Расплавы, № 3, 2010, С. 86–96.

14. Потапов А.М., Кумков С.И., Sato Y. Обработка экспериментальных данных по вязкости при одностороннем смещении ошибок измерения // Расплавы, № 3, 2010, С. 55–70.

15. Кумков С.И. Разработка совместного Российского – ISO стандарта (методики) обработки измерительной информации в условиях неопределённости ошибок измерений и малого числа наблюдений (на основе методов интервального анализа) // Доклады Всероссийского (с международным участием) Совещания по интервальному анализу и его приложениям "ИНТЕРВАЛ-06", 1–4 июля 2006 г., Санкт–Петербург, Санкт–Петербургский Государственный Университет, 2006, стр. 63–67, и сайт <http://www.ict.nac.ru/interval/Conferences/Interval-06>

Содержание настоящего сообщения:

На основе методов интервального анализа [7,9–11] рассматривается подход к обработке выборки зашумленных экспериментальных данных при малом числе значений основного аргумента и многократных измерениях в подвыборках для каждого его значения.

Задан теоретический вид исследуемой зависимости с вектором её параметров. Значения аргумента известны точно, а замеры в подвыборках содержат как обычные приборные погрешности измерений, так и хаотические искажения неизвестного знака и величины. Вероятностные характеристики обеих компонент погрешностей не являются гауссовскими либо неизвестны, неопределённой является также величина ограничения на максимальное значение суммарных погрешностей. Поэтому невозможно обосновать применение стандартных (ГОСТ) процедур обработки, опирающихся на статистические методы и предположения: о представительности (большое число значений основного аргумента), о нормальности распределения погрешностей измерения, о статистической однородности погрешности многократных измерений и о несмещённости замеров.

Исходные данные [1]

Заданы:

– зашумленная выборка замеров электрохимического процесса с упорядоченными по возрастанию значениями аргумента T_i и многократными измерениями $n_i = 1, N_i$ при каждом его i -ом значении

$$\{T_i, \{E_{i,n_i}\}\}, T_1 < T_2 < T_3, i = 1, 3, n_i = 1, N_i; \quad (1)$$

– описывающая функция – линейная зависимость

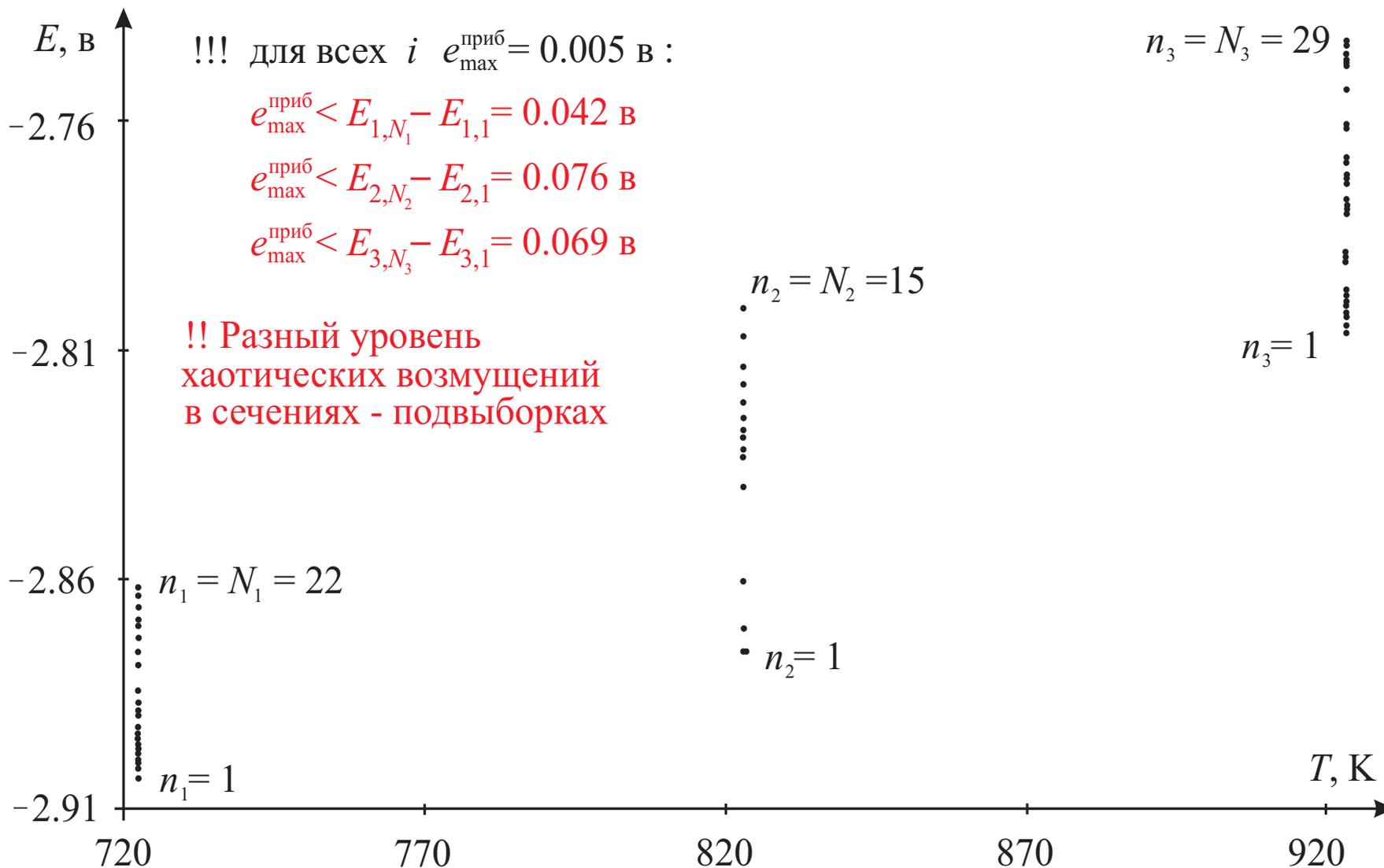
$$E(T) = BT + A; \quad B, A \text{ – параметры}; \quad (2)$$

– структура зашумленных замеров имеет вид

$$E_{i,n_i} = E_{i,n_i}^{\text{ИСТ}} + e_{i,n_i}, \quad |e_{i,n_i}| \leq e_{\text{max}}, \quad e_{i,n_i} = e_{i,n_i}^{\text{приб}} + \chi_{i,n_i}, \quad (3)$$

где $E_{i,n_i}^{\text{ИСТ}}$ – неизвестное истинное значение потенциала, измеряемого при значении аргумента T_i в n_i -ом наблюдении, $n_i = 1, N_i$; замеры $\{E_{i,n_i}\}$ в подвыборках занумерованы по возрастанию; e_{i,n_i} – суммарная погрешность измерения с неизвестной величиной ограничения e_{max} , $e_{i,n_i}^{\text{приб}}$ – приборная погрешность, χ_{i,n_i} – хаотическая компонента суммарной погрешности.

Исходные данные [1]. Короткая выборка. Многократные измерения. !!! Хаотические искажения



Классическая инженерная задача аппроксимации и оценивания параметров исследуемой зависимости по [4–6]

Условия применения стандартных статистических процедур:

- выборка представительна, т.е. её длина N "достаточно велика";
- погрешность измерения e имеет вероятностный характер;
- ее распределение является нормальным и несмещённым, $m_e = 0$;
- значения основного аргумента точные.

Квадратичный критерий невязки замеров заданной выборки $\{E_{i,n_i}\}$ относительно средней аппроксимирующей зависимости $E(T_i, \Gamma)$ (для соответствующего значения Γ вектора параметров)

$$\begin{aligned} J^{KB}(\Gamma) &= \sum_{i=1,3;n_i=1,N_i} (\Delta_{i,n_i})^2, \\ \Delta_{i,n_i} &= E_{i,n_i} - E(T_i, \Gamma). \end{aligned} \quad (4)$$

Решение:

$$\begin{aligned} \Gamma^{KB} : \arg\{\min_{\Gamma} J^{KB}(\Gamma)\}, \text{ и при заданной } P_{\text{дов}} = 0.97 \text{ находятся:} \\ \text{доверительная точность оценки параметров } \pm \delta_{\Gamma} \text{ и} \\ \text{разброс значений зависимости } \pm 3\sigma^{\text{ЭМП}}. \end{aligned} \quad (5)$$

Глобальная идеология статистических подходов [4–6]

Анализ рассеяния замеров относительно среднего значения (или средней зависимости).

Объявление замера выбросом (промахом) и его отсев по величине отклонения с заданной доверительной вероятностью при допущении некоторого закона распределения вероятностей их рассеяния.

Условия практических экспериментов [1–3]

На практике:

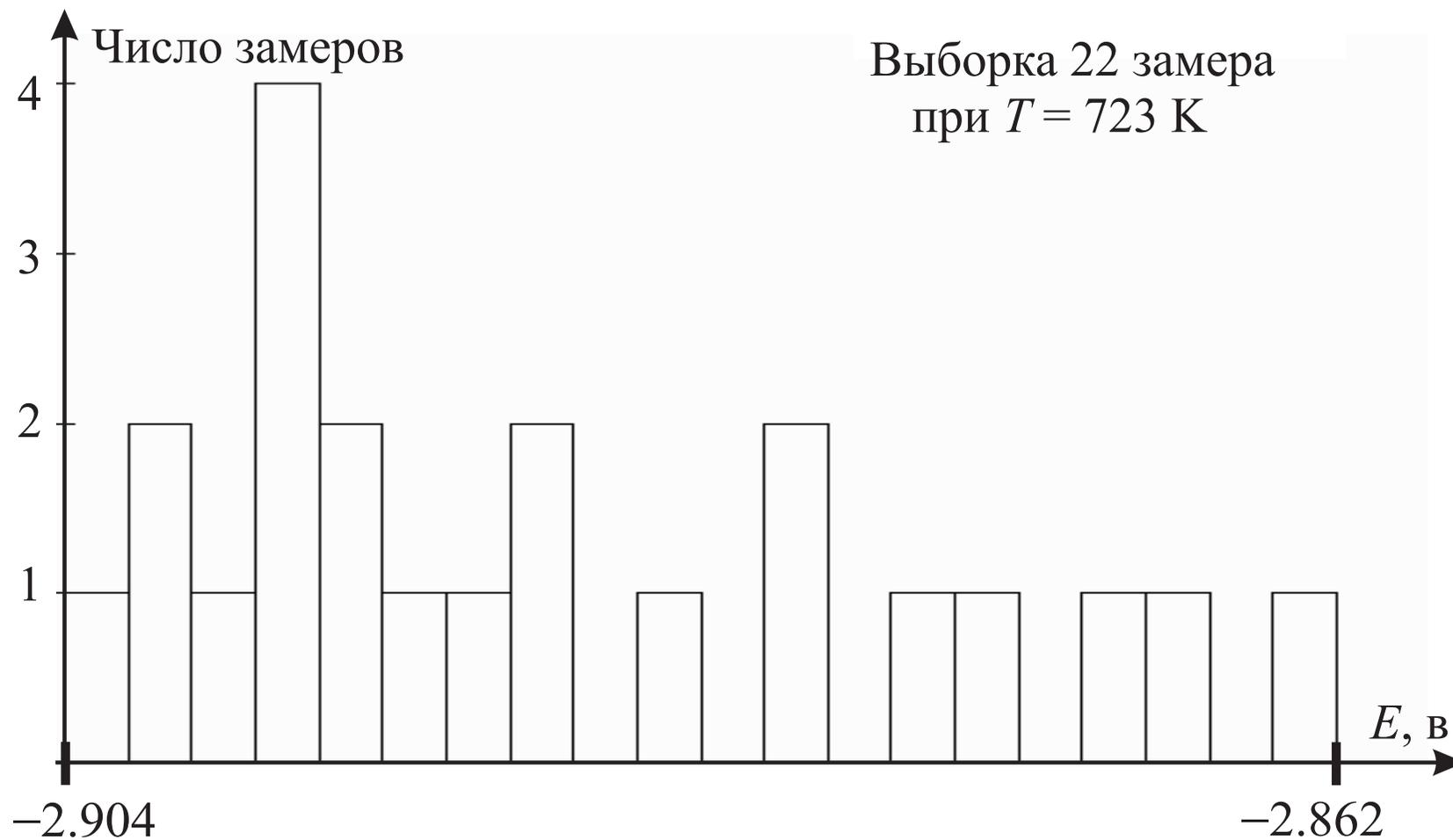
- выборка короткая, например, всего 3-4 замера [1–3];
- суммарная погрешность измерений имеет структуру

$$e = e^{\text{приб}} + \chi, \quad |e| \leq e_{\text{max}}, \quad (6)$$

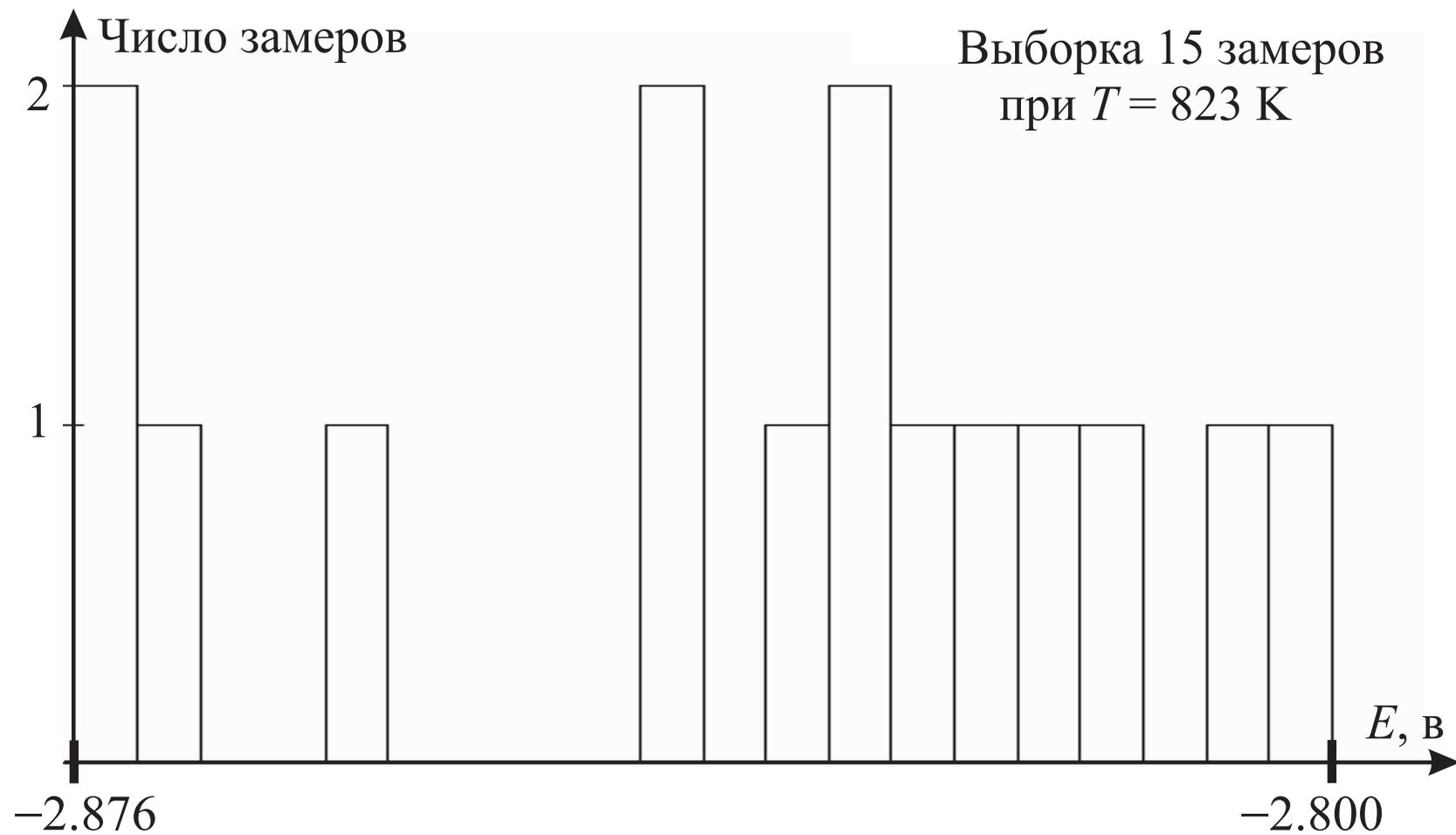
где $e^{\text{приб}}$ – приборная погрешность измерений; χ – неконтролируемое (хаотическое) искажение замера; e_{max} – ограничение на максимальное значение суммарной погрешности; при этом:

- вероятностные характеристики обеих компонент и их характер полностью неизвестны, распределения не являются нормальными;
- величина ограничения e_{max} известна лишь приближённо из собственных *паспортных данных* измерительной системы; как правило, эта оценка является заниженной;
- измерение E_{i,n_i} может быть не прямым, а *косвенным*, т.е. величины E_{i,n_i} в (1) содержат дополнительные неконтролируемые погрешности, появляющиеся в результате некоторых вычислений над первичными измерениями;
- погрешности могут присутствовать как в значениях замеров, так и в значениях аргумента.

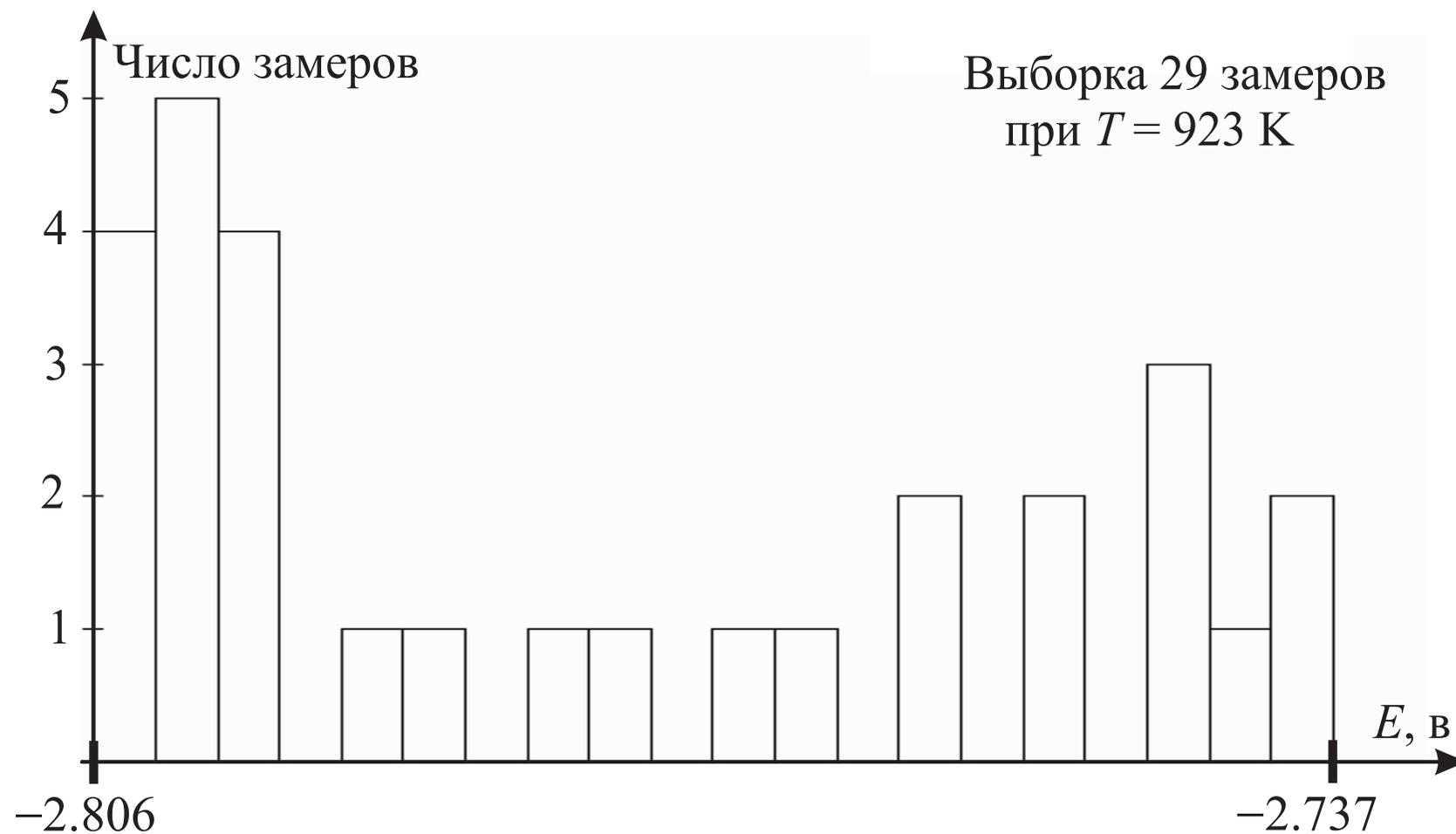
Типичная гистограмма рассеяния замеров [1]



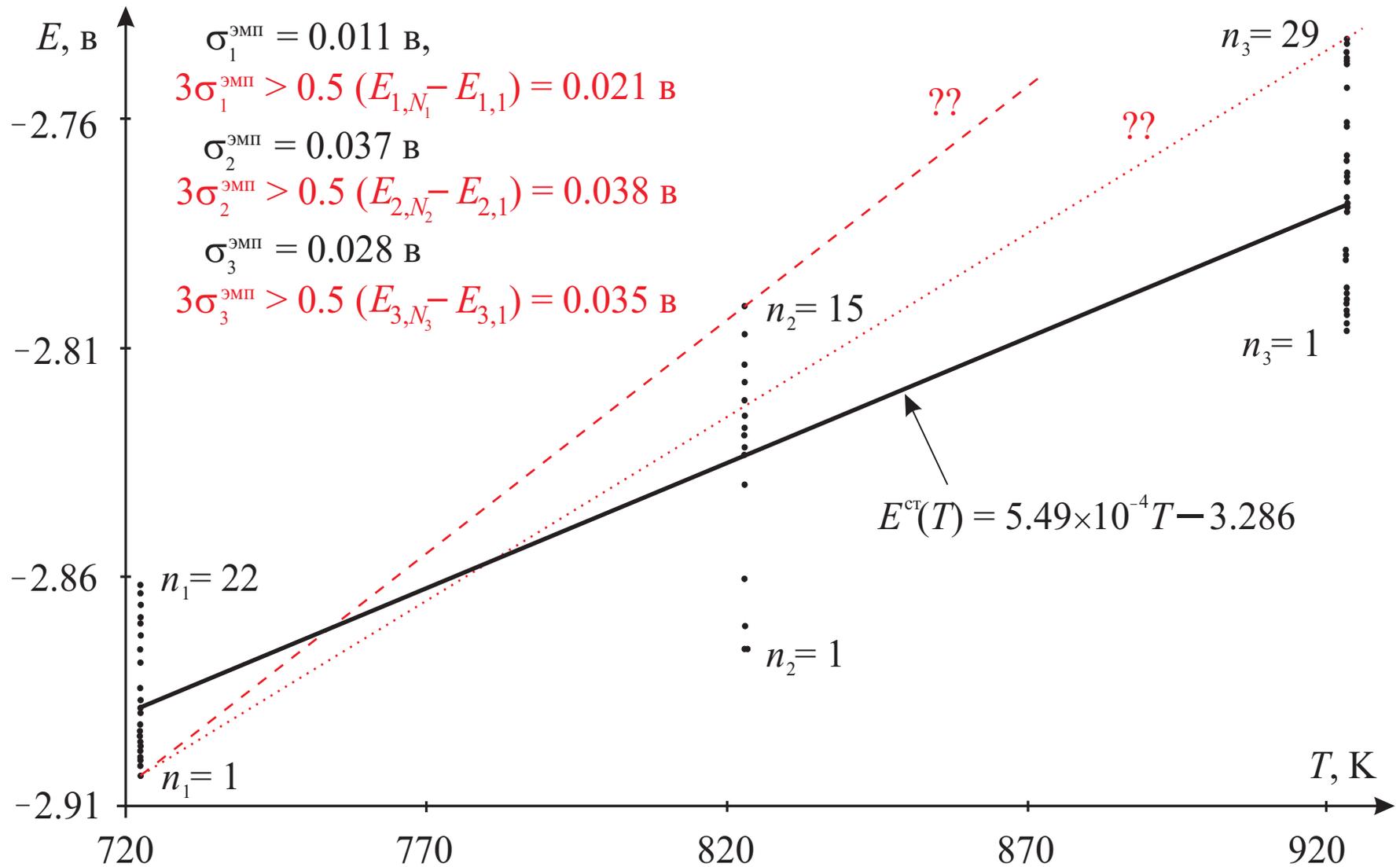
Типичная гистограмма рассеяния замеров [1]



Типичная гистограмма рассеяния замеров [1]



Формальное применение стандартного подхода:



Интервальный подход [8,12–15]. Определения.

Для замеров выборки (1) при **заданном**, приборном уровне $e_{\max}^{\text{приб}}$ ограничении на погрешность измерений строится набор *интервалов неопределённости* (ИНЗ) с соответствующими границами [8,13–16]

$$H_{i,n_i} = [\underline{h}_{i,n_i}, \bar{h}_{i,n_i}] : \underline{h}_{i,n_i} = E_{i,n_i} - e_{\max}^{\text{приб}}, \bar{h}_{i,n_i} = E_{i,n_i} + e_{\max}^{\text{приб}}. \quad (7)$$

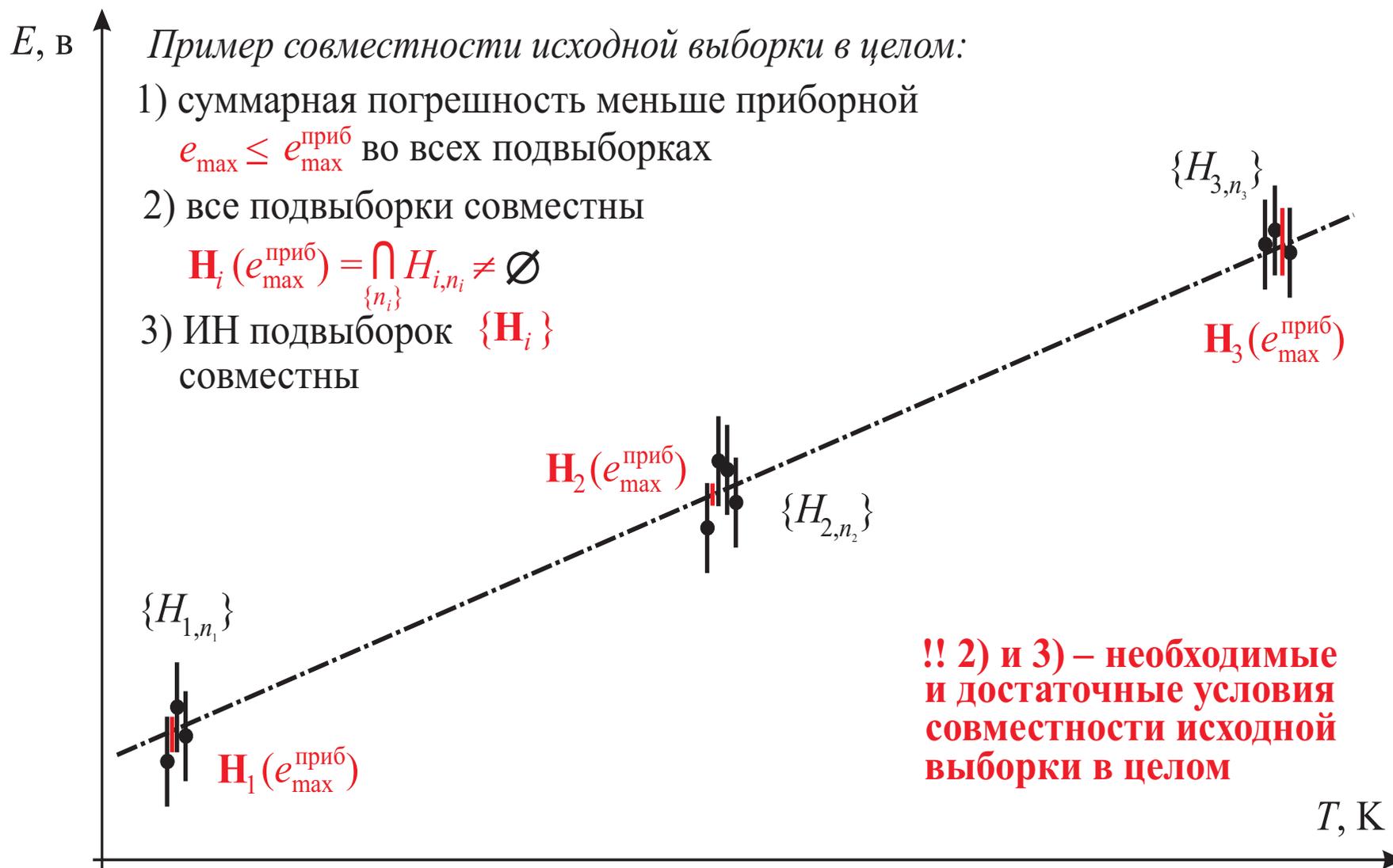
Выборка полагается **совместной**, если через **все** ИНЗ замеров можно провести хотя бы одну зависимость (описывающую функцию) заданного вида. Каждая такая зависимость полагается **допустимой**.

Совокупность всех допустимых зависимостей (т.е. проходящих через совместный набор ИНЗ замеров) полагается **трубкой** допустимых зависимостей данного набора.

Совместному набору допустимых зависимостей ставится в соответствие множество допустимых значений оцениваемых параметров в пространстве параметров. Такое множество назовём **информационным множеством параметров** $I(B, A, e_{\max})$. В рассматриваемом варианте примера (1) с многократными наблюдениями это результирующее (т.е., соответствующее всей выборке замеров) множество составляется из всех пар (B, A) значений параметров, для которых

$$I(B, A, e_{\max}^{\text{приб}}) = \left\{ B, A : \underline{h}_{i,n_i} \leq BT_i + A \leq \bar{h}_{i,n_i}, \text{ для всех } i \text{ и } n_i \right\}. \quad (8)$$

Ситуация при совместности исходной выборки



I. ПОСТРОЕНИЕ РЕЗУЛЬТИРУЮЩЕГО
ИНФОРМАЦИОННОГО МНОЖЕСТВА.
ОСНОВНЫЕ ПРОЦЕДУРЫ.
ПРОБЛЕМЫ

Ситуация при наличии хаотических искажений в замерах

ШАГ 1. Выделение совместных сочетаний последовательных (по значению основного аргумента) замеров.

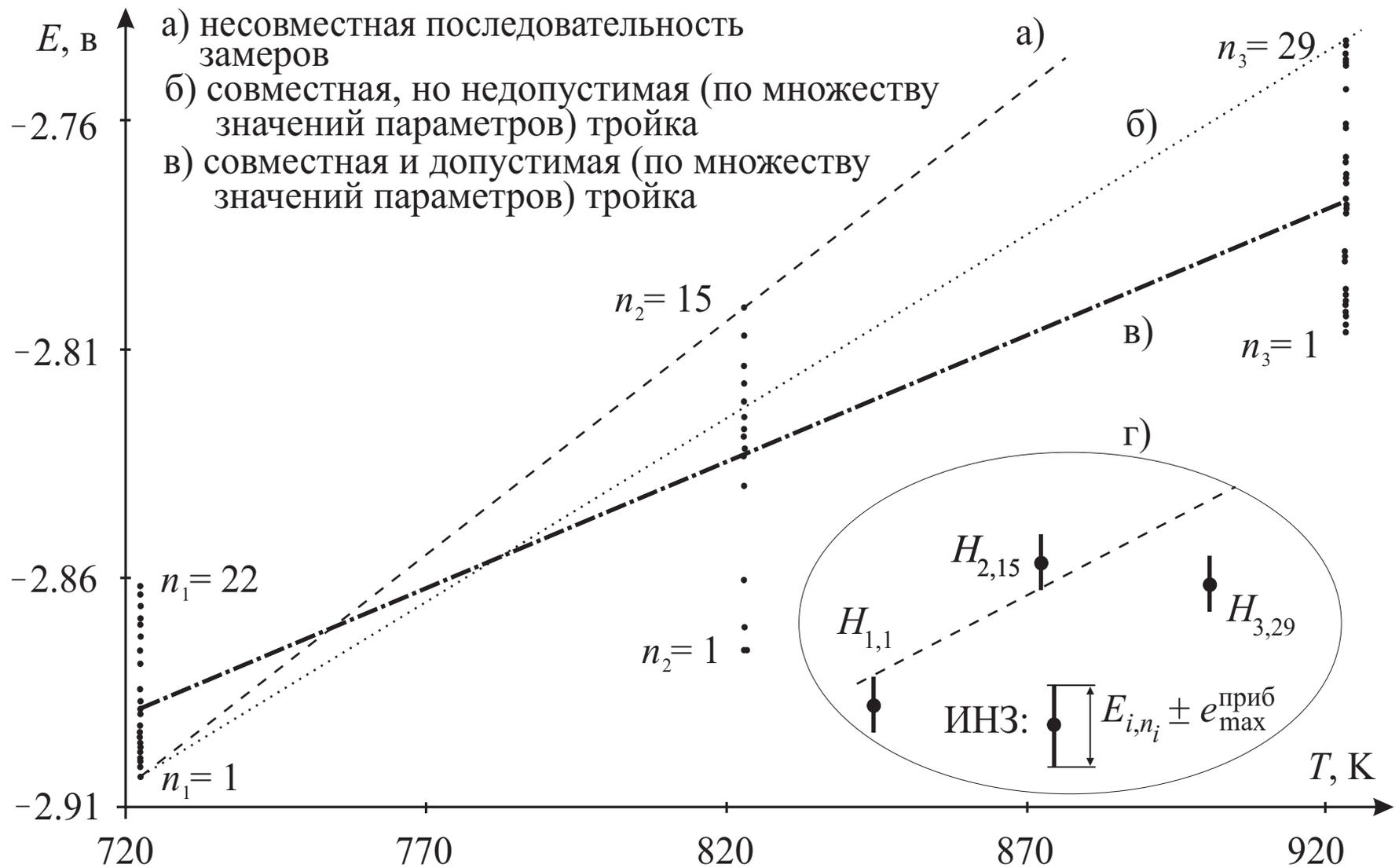
Для замеров выборки (1) рассматривается [1] набор троек последовательных замеров для значений аргумента T_1 , T_2 и T_3 с их интервалами неопределённости

$$\{H_{1,k}, H_{2,l}, H_{3,m}\}, \quad k = 1, N_1, \quad l = 1, N_2, \quad m = 1, N_3. \quad (9)$$

Каждая тройка (9) полагается *совместной*, если через *все* эти три ИНЗ можно провести хотя бы одну зависимость (описывающую функцию) заданного вида. Каждая такая зависимость полагается *допустимой*.

Каждой *совместной* тройке ставится в соответствие информационное множество $I_{k,l,m}(B, A, e_{\max}^{\text{приб}})$ допустимых значений её параметров. Для несовместной тройки такое множество пусто.

Совместные и несовместные тройки ИНЗ



Информационное множество совместной тройки

Задача расчёта информационного множества

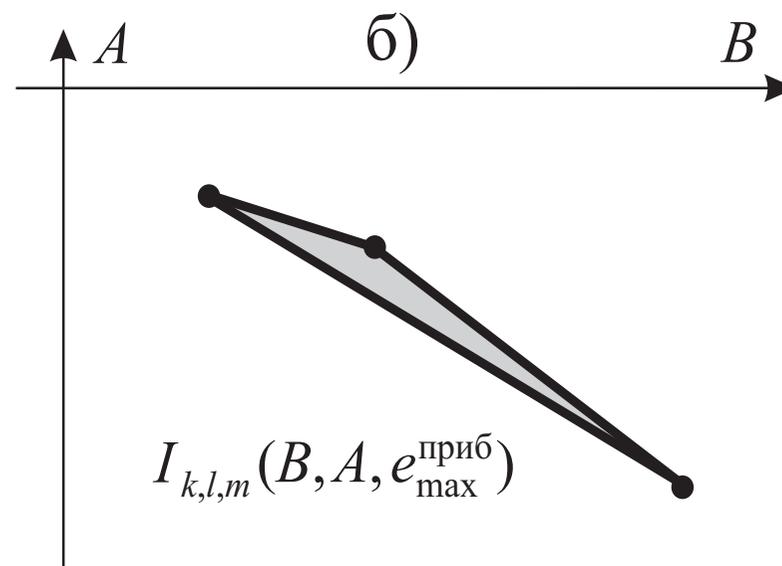
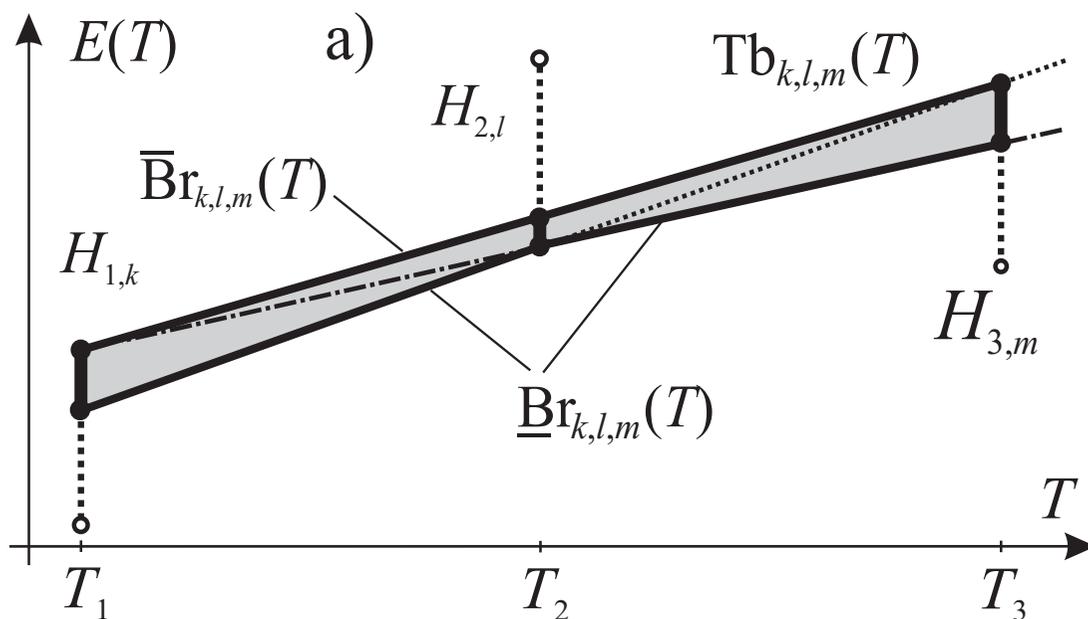
$I_{k,l,m}(B, A, e_{\max}^{\text{приб}})$ совместной тройки сводится [8,12–15] к задаче решения системы трёх интервальных неравенств

$$\begin{aligned} I_{k,l,m}(B, A, e_{\max}^{\text{приб}}) = \{ B, A : \\ \underline{h}_{1,k} \leq BT_1 + A \leq \bar{h}_{1,k}, \\ \underline{h}_{2,l} \leq BT_2 + A \leq \bar{h}_{2,l}, \\ \underline{h}_{3,m} \leq BT_3 + A \leq \bar{h}_{3,m}. \end{aligned} \quad (10)$$

Информационное множество совместной тройки:

а) интервалы неопределённости замеров и трубка $Tb_{k,l,m}$ допустимых зависимостей с верхней $\overline{Br}_{k,l,m}$ и нижней $\underline{Br}_{k,l,m}$ границами; отсекаемые (пунктир) и остающиеся (жирные вертикальные отрезки) части ИНЗ;

б) информационное множество параметров совместной тройки.



Учет априорной информации о допустимых значениях параметров

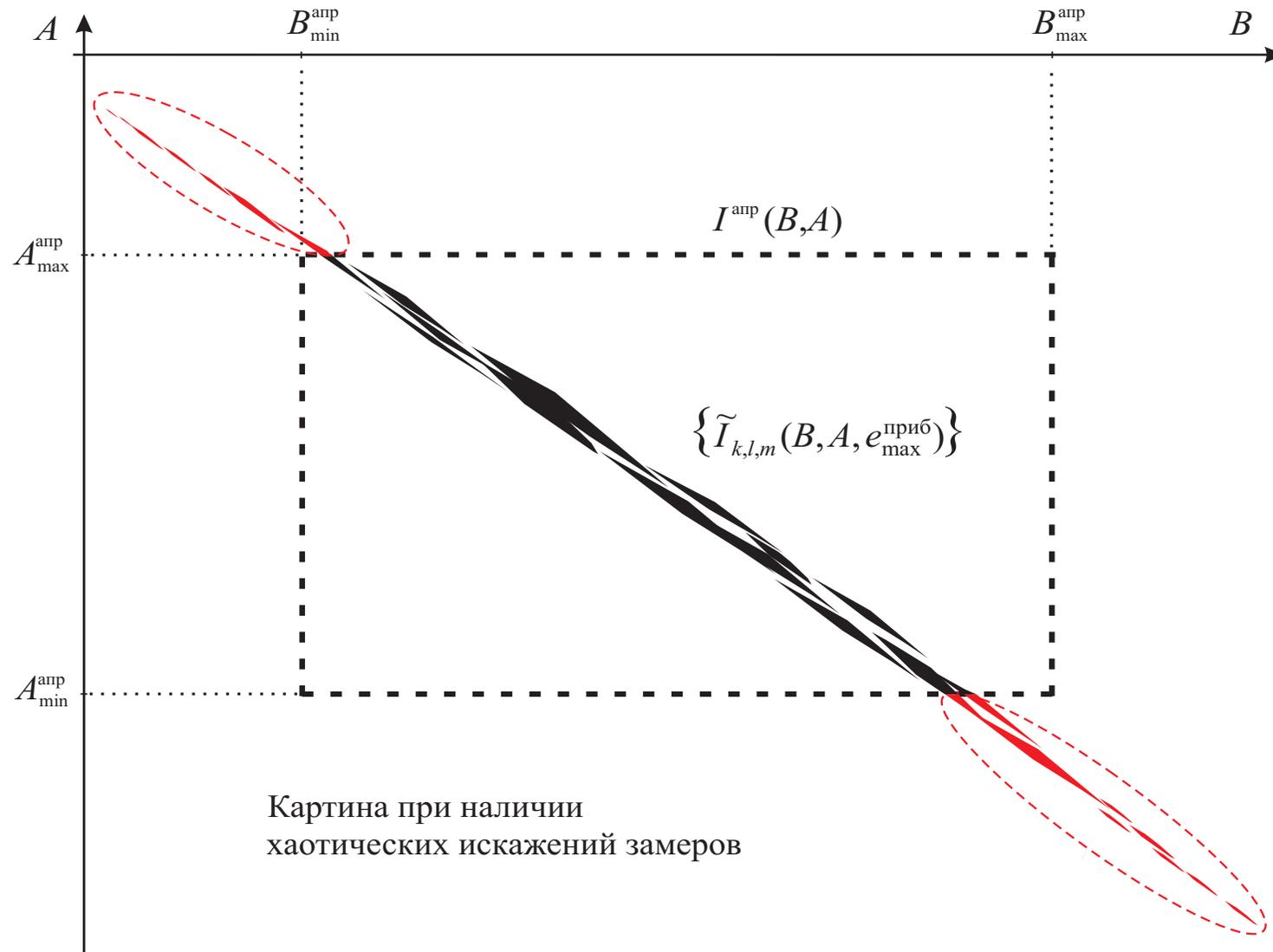
ШАГ 2. Учет априорного множества $I^{\text{апр}}(B, A)$ параметров и коррекция набора информационных множеств $\{I_{k,l,m}(B, A, e_{\text{max}}^{\text{приб}})\}$ допустимых значений параметров

$$\tilde{I}_{k,l,m}(B, A, e_{\text{max}}^{\text{приб}}) = I^{\text{апр}}(B, A) \cap I_{k,l,m}(B, A, e_{\text{max}}^{\text{приб}}),$$

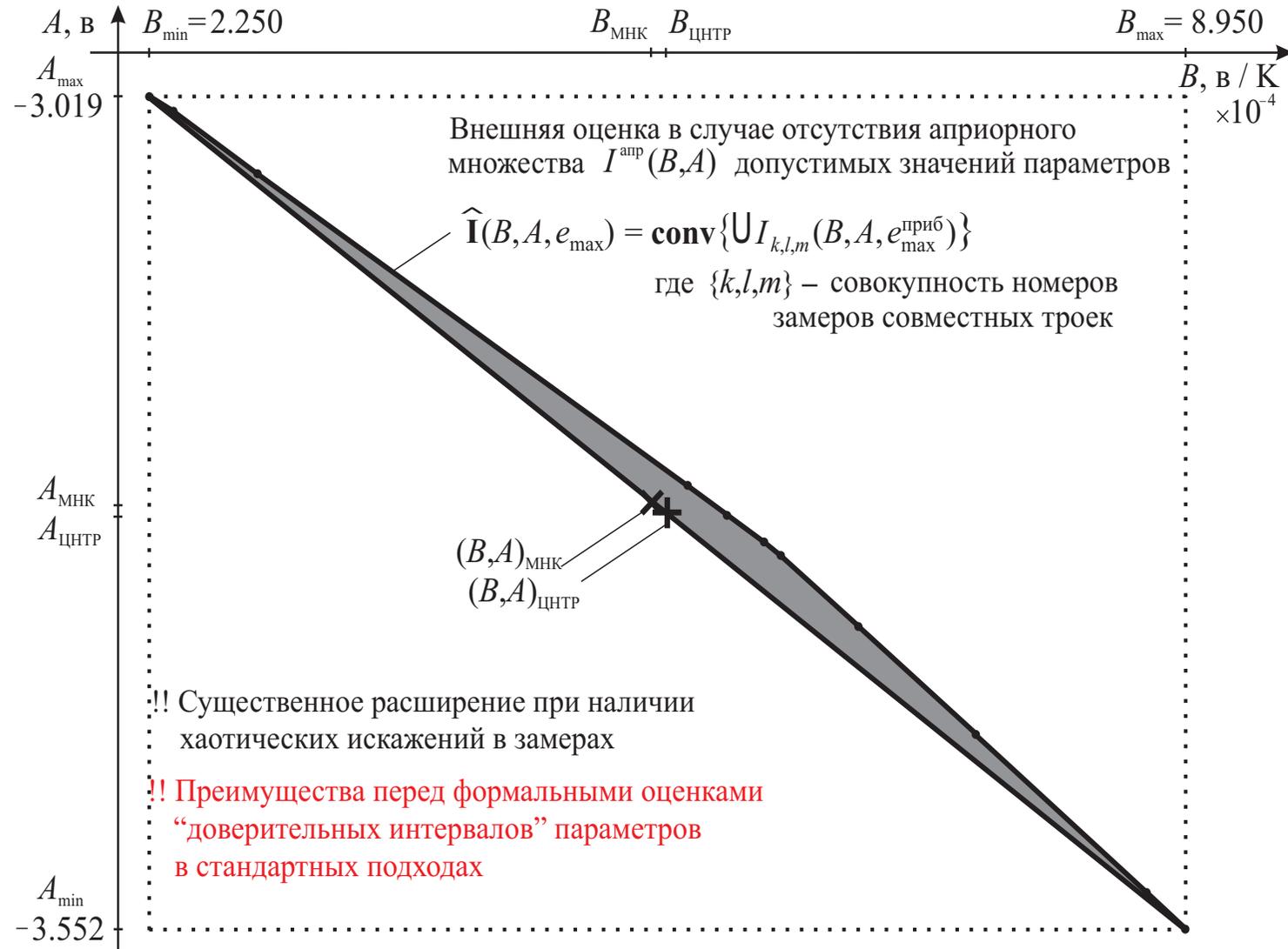
$\{k, l, m\}$ – номера замеров совместных троек, (11)

$$k \in [1, N_1], \quad l \in [1, N_2], \quad m \in [1, N_3].$$

Уточнение набора информационных множеств совместных троек [1]



Внешняя оценка результирующего информационно-го множества [1]



Идеология применения интервального анализа

Выполняется анализ допустимости не **отдельных** замеров, а более глубокого свойства – **совместности их последовательностей** в соответствии с заданным типом зависимости.

В отличие от стандартных статистических подходов, правило отсева опирается на использование только **допустимых информационных множеств параметров** каждой последовательности. Никакие допущения о неизвестных распределениях вероятностей не требуются.

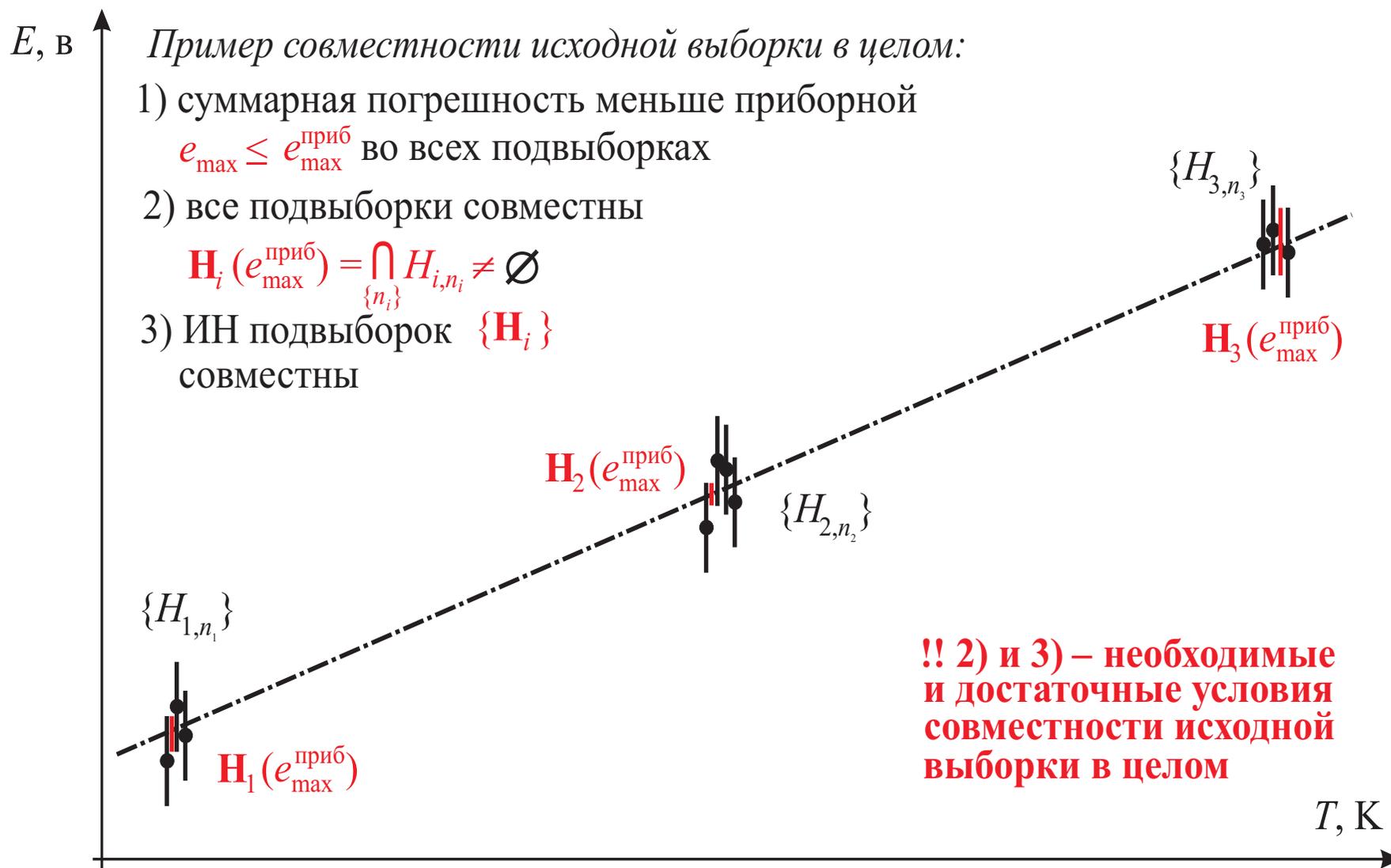
В дальнейшем анализе используются **только замеры, дающие совместные последовательности** (в рассматриваемом примере – тройки).

ПРОБЛЕМЫ РЕАЛИЗАЦИИ ПОДХОДА

1. Неопределённость **фактического** уровня суммарной погрешности в замерах.
2. Возможность отсутствия, “неправильности” или грубости задания априорного множества.
3. Завышенность внешней оценки результирующего информационного множества из-за хаотических искажений замеров.
4. Повышенный объем вычислений из-за комбинаторного перебора последовательностей (в [1] – троек) замеров.

II. ПРЕДЛАГАЕМАЯ МЕТОДИКА
С ОПРЕДЕЛЕНИЕМ
ФАКТИЧЕСКОГО УРОВНЯ
СУММАРНОЙ ПОГРЕШНОСТИ
И ВАРИАЦИЕЙ ЕЁ ОГРАНИЧЕНИЯ

Случай А: совместность исходной выборки в целом



Фактический уровень суммарной погрешности меньше приборной во всех подвыборках.

Случай Б: несовместность исходной выборки в целом

$E, \text{ В}$ ↑ *Пример несовместности исходной выборки в целом:*

1) суммарная погрешность в подвыборках
больше приборной $e_{\max} > e_{\max}^{\text{приб}}$

2) все подвыборки совместны

$$\mathbf{H}_i(e_{\max}^{\text{приб}}) = \bigcap_{\{n_i\}} H_{i,n_i} \neq \emptyset$$

3) ИНЗ подвыборок $\{\mathbf{H}_i\}$ несовместны

4) нет ни одной совместной тройки

$\{H_{1,n_1}\}$



$\{H_{2,n_2}\}$



$\{H_{3,n_3}\}$



$T, \text{ К}$ →

Случай В: несовместность исходной выборки в целом

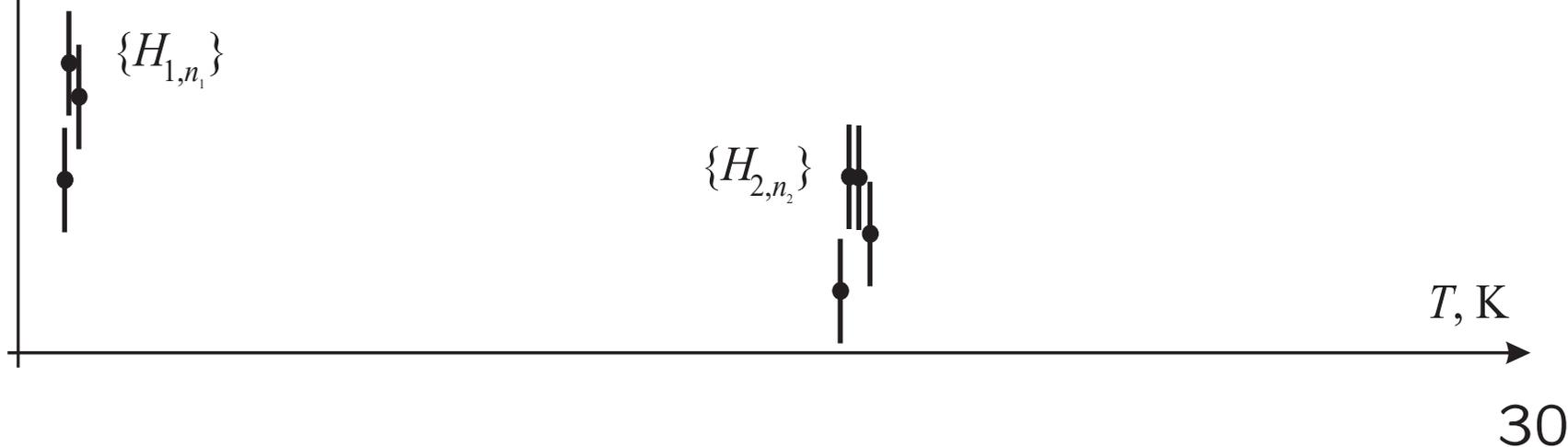
$E, \text{ В}$ ↑ *Пример несовместности исходной выборки в целом:*

1) суммарная погрешность в подвыборках
больше приборной $e_{\max} > e_{\max}^{\text{приб}}$

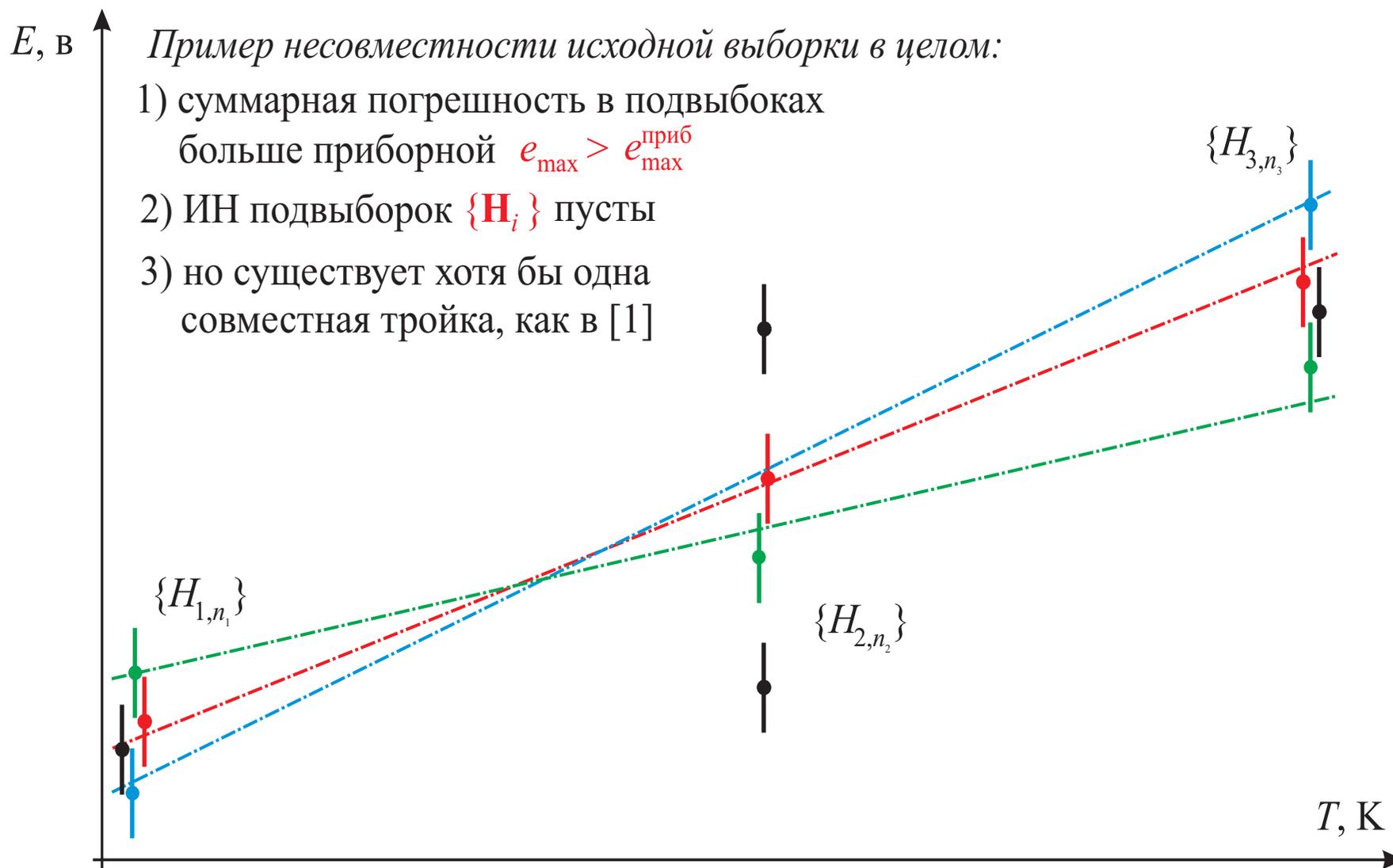
2) хотя бы одна подвыборка несовместна

$$H_i(e_{\max}^{\text{приб}}) = \bigcap_{\{n_i\}} H_{i,n_i} = \emptyset$$

3) нет ни одной совместной тройки



Случай Г: несовместность исходной выборки в целом



ПРОЦЕДУРЫ ПРЕДЛАГАЕМОЙ МЕТОДИКИ

ПРОЦЕДУРА 1. Определение рабочих номеров минимального \underline{n}_i и максимального \bar{n}_i замеров в каждой подвыборке

Случай А, все подвыборки совместны и их ИН \mathbf{H}_i совместны:

$$\underline{n}_i = n_{i,1}, \bar{n}_i = N_i.$$

Случай Б, все подвыборки совместны, но их ИН \mathbf{H}_i несовместны и нет ни одной совместной тройки:

$$\underline{n}_i = n_{i,1}, \bar{n}_i = N_i, \text{ как для непустых, так и для пустых ИН } \mathbf{H}_i \text{ подвыборок.}$$

Случай В, хотя бы одна подвыборка несовместна и нет ни одной совместной тройки: $\underline{n}_i = n_{i,1}, \bar{n}_i = N_i$, как для непустых, так и для пустых ИН \mathbf{H}_i подвыборок.

Случай Г, ни одна подвыборка не совместна, но есть хотя бы одна совместная тройка:

$\underline{n}_1 = \min\{k\}, \bar{n}_1 = \max\{k\}$, где $\{k\}$ – совокупность номеров ИНЗ первого сечения, дающих непустые тройки,

$\underline{n}_2 = \min\{l\}, \bar{n}_2 = \max\{l\}$, где $\{l\}$ – совокупность номеров ИНЗ второго сечения, дающих непустые тройки,

$\underline{n}_3 = \min\{m\}, \bar{n}_3 = \max\{m\}$, где $\{m\}$ – совокупность номеров ИНЗ третьего сечения, дающих непустые тройки.

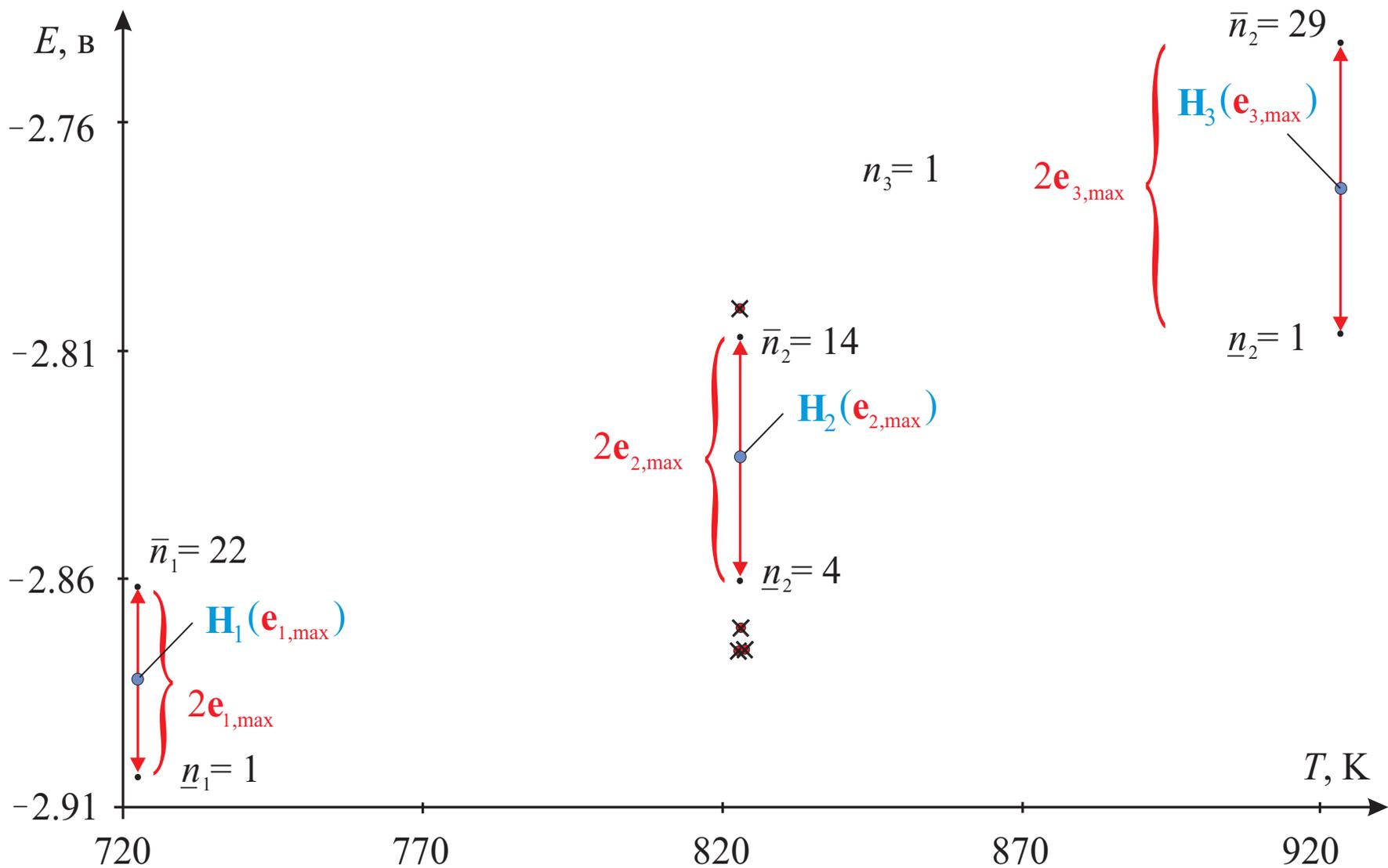
ПРОЦЕДУРЫ ПРЕДЛАГАЕМОЙ МЕТОДИКИ

ПРОЦЕДУРА 2. Оценка снизу фактического уровня суммарной погрешности измерений в каждом сечении выборки:

Расчёт минимального уровня $e_{i,\max}$, при котором **минимальные** результирующие интервалы неопределённости \mathbf{H}_i в каждой подвыборке замеров (в каждом сечении выборки) являются **точечными**. Эти уровни равны полуразмахам разброса замеров с рабочими номерами \underline{n}_i и \bar{n}_i каждой подвыборки

$$e_{i,\max} = 0.5(E_{i,\bar{n}_i} - E_{i,\underline{n}_i}).$$

Результат Процедур 1 и 2



ПРОЦЕДУРЫ ПРЕДЛАГАЕМОЙ МЕТОДИКИ

ПРОЦЕДУРА 3. Введение вариации уровня ограничения

$$e_{i,\max}^{\text{вар}} = \delta^{\text{вар}} e_{i,\max}$$

суммарной погрешности в каждом сечении, где $\delta^{\text{вар}} > 1$ – вариация уровня увеличения суммарной погрешности.

ПРОЦЕДУРЫ ПРЕДЛАГАЕМОЙ МЕТОДИКИ

ПРОЦЕДУРА 4. Вариация уровня ограничения.

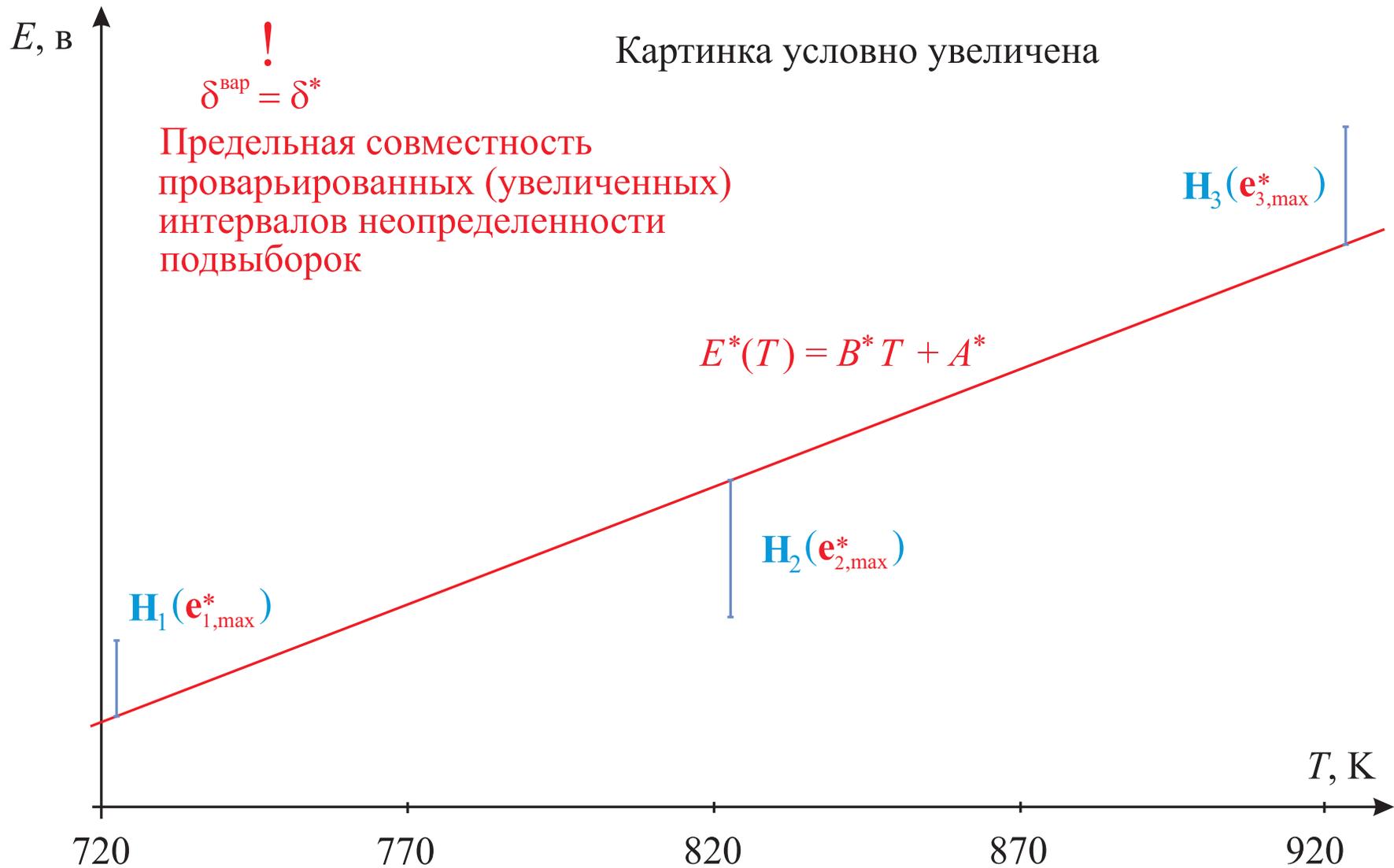
Вариация выполняется от нуля в сторону увеличения до предельного значения коэффициента δ^* с соответствующим увеличением ограничений

$$\mathbf{e}_{i,\max}^* = \delta^* \mathbf{e}_{i,\max}, \quad \delta^* > 1,$$

при которых **телесные** (уже не точечные) интервалы неопределённости $\{\mathbf{H}_i(\mathbf{e}_{i,\max}^*)\}$ всех сечений становятся **предельно-совместными**, т.е. через них можно провести единственную допустимую зависимость

$$E^*(T) = B^*T + A^*.$$

Результат Процедур 3 и 4



ПРОЦЕДУРЫ ПРЕДЛАГАЕМОЙ МЕТОДИКИ

ПРОЦЕДУРА 5. Задание **доверительного** уровня суммарной погрешности:

На основе предельных величин ограничений $\{e_{i,\max}^*\}$ из инженерных соображений рассчитываются **доверительные** уровни $\{e_{i,\max}^{\text{ДОВ}}\}$ ограничений на суммарную погрешность в каждом сечении

$$e_{i,\max}^{\text{ДОВ}} = (1 + \mu^{\text{ДОВ}}) e_{i,\max}^*, \quad \mu^{\text{ДОВ}} \approx 5 - 10\%.$$

ПРОЦЕДУРЫ ПРЕДЛАГАЕМОЙ МЕТОДИКИ

ПРОЦЕДУРА 6. Расчёт выходных данных обработки.

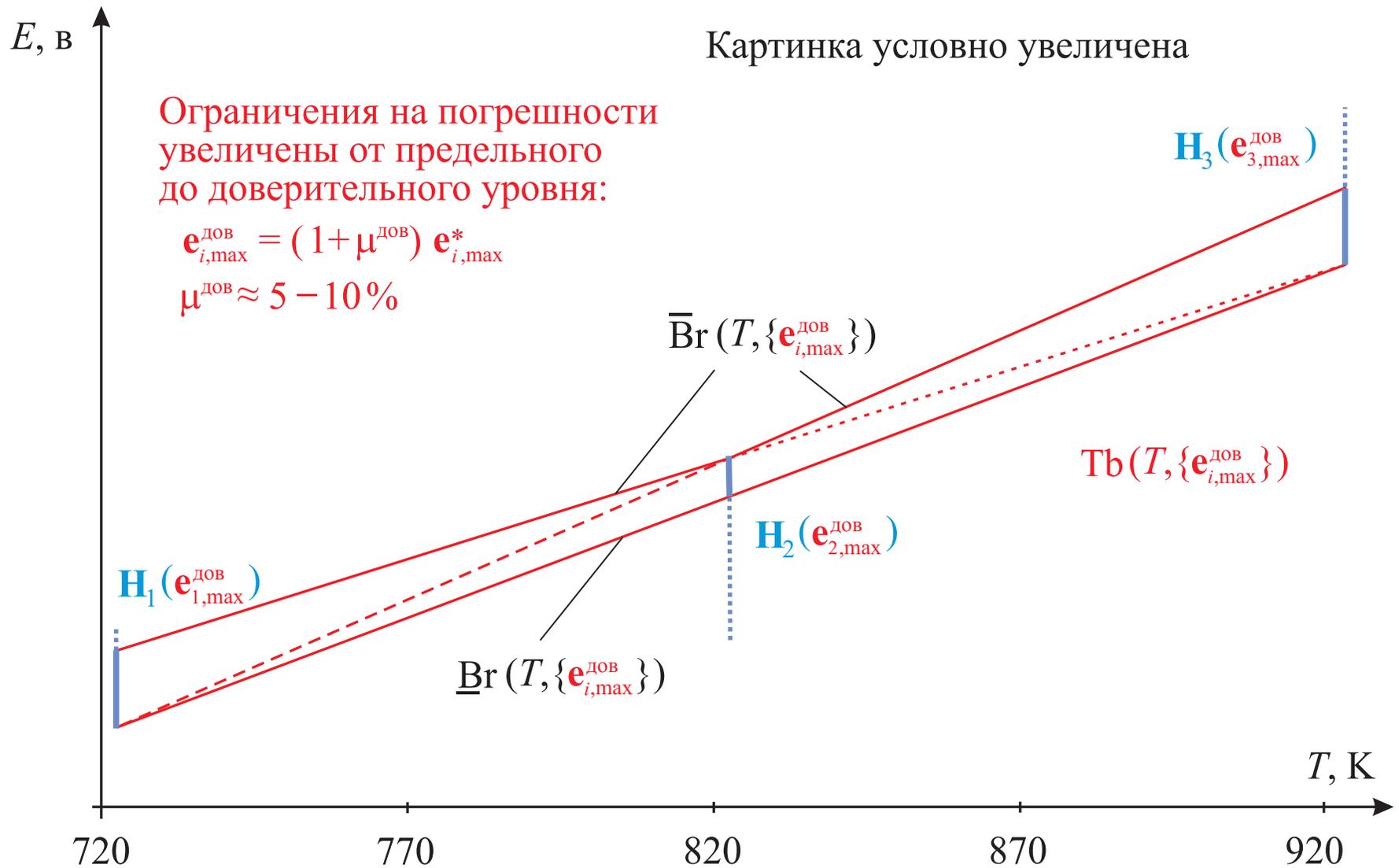
Расчёт результирующего **доверительного** информационного множества параметров

$$\mathbf{I}(B, A, \{ e_{i,\max}^{\text{ДОВ}} \}).$$

Расчёт **доверительной** трубки допустимых зависимостей

$$\text{Tb}(T, \{ e_{i,\max}^{\text{ДОВ}} \}).$$

Результат Процедур 5 и 6



ЗАКЛЮЧЕНИЕ

В указанных условиях неопределённости вероятностных характеристик как приборных погрешностей измерения, так и хаотических искажений замеров гарантируется:

1. **СОСТОЯТЕЛЬНОСТЬ** методики, т.е. сходимость результатов оценивания к истинным величинам параметров при уменьшении к нулю хаотических искажений и приборных погрешностей измерения.
2. Получение **ГАРАНТИРОВАННЫХ** оценок множества допустимых значений параметров и трубки допустимых зависимостей исследуемого типа в зависимости от **ДОВЕРИТЕЛЬНОГО** уровня суммарных погрешностей измерений в каждом сечении выборки замеров (при многократных измерениях).

СПАСИБО ЗА ВНИМАНИЕ.

Исходные данные [14]. Хаотические искажения

