

**МЕТРОЛОГИЧЕСКИЕ ОЦЕНКИ НЕОПРЕДЕЛЕННОСТИ  
В КАЧЕСТВЕННОМ АНАЛИЗЕ**

**В.И.Вершинин**

**Россия, Омск, Омский государственный университет,  
кафедра аналитической химии**

**vershin @ univer.omsk.su**

## Предпосылки повышенного внимания к метрологии качественного анализа на рубеже XX-XXI веков

- 1) Резкий рост числа объектов обнаружения, особенно органических. Необходимость надежного экспрессного обнаружения супертоксиантов, лекарственных препаратов, допинга, наркотиков, взрывчатых веществ и т.п., в том числе с применением тест-методов.
- 2) Развитие инструментальных методов, ведущее к ситуациям, когда ошибки анализа обусловлены не процессом измерения сигнала или градуировки, а неверной идентификацией компонентов пробы.
- 3) Компьютеризация анализа, создание больших баз данных по свойствам веществ и, на их основе, развитие **систем компьютерной идентификации (СКИ)**.

# Методы анализа, в которых применяют системы компьютерной идентификации (СКИ)

- Газовая хроматография      Анализ бензинов, растворителей, пищевых продуктов и др.
- Жидкостная хроматография      Определение пестицидов, наркотиков, анализ лекарственных препаратов и др.
- Хромато-масс-спектрометрия      Анализ пищевых продуктов и др.
- Масс-спектрометрия, ЯМР      Идентификация продуктов лабораторного органического синтеза
- Спектрофлуориметрия      Обнаружение ПАУ и нитрозаминов
- Атомно-эмиссионный спектральный (ICP)      Анализ сточных вод и др.
- ИК-спектрометрия      Анализ лекарственных препаратов и др.
- Рентгенофазовый      Анализ минерального сырья и др.
- Рентгенофлуоресцентный      Анализ минерального сырья и др.

## Пример выдачи на печать результатов работы системы компьютерной идентификации (СКИ) при хроматографическом анализе бензина по ASTM 5134

Peak	t	I	I <sub>x</sub>	Name	%
37	33,975	741,1	741,3	<b>3,3 - dimethylhexane</b>	<b>0,109</b>
38	35,718	755,3	755,3	<b>Toluene</b>	<b>11,066</b>
39	35,944	756,4	?	<b>unidentified</b>	<b>0,018</b>
40	36,367	760,6	760,2	<b>2,3 - dimethylhexane</b>	<b>0,319</b>
41	36,541	762,0	762,1	<b>3- ethyl - 2 - methylpentane</b>	<b>0,078</b>

t – время удерживания, минуты

I – индекс Ковача для опознаваемого пика пробы,

I<sub>x</sub> - индекс Ковача для пика X<sub>i</sub> в БД,

d - критерий совпадения пиков, здесь - 0,5 единицы индекса

% - содержание компонента ( в % масс.).

## Некоторые отечественные публикации в области метрологии качественного анализа

- **1955** - Комарь Н.П. Основы качественного химического анализа. Харьков
- **1976** - Бугаевский А.А. и др. Зав.лаборатория. Т.42, №1, с.68.
- **1987** - Вершинин В.И., Топчий В.А. и др. ЖАХ. Т.42, с.837-845.
- **1995** – Зенкевич И.Г. и др. ЖАХ. Т.50, № 2, с.118.
- **1999** – Мильман Б.Л., Конопелько Л.А. Зав.лаборатория. 1999. Т.65, № 12
- **2000** - Milman B.L., Konopelko L.A. Fresenius J.Anal.Chem. V.367, p.621-28.
- **2002** - Вершинин В.И., Дерендяев Б.Г., Лебедев К.С. Компьютерная идентификация органических соединений. М., Академкнига. 197 с.
- **2003** – Решетняк Е.А. и др. Вестник ХНУ. Вып.596, № 10. с. 90
- **2004** – Островская В.М., Решетняк Е.А. и др. ЖАХ, т.59, №10, с.1101.
- **2004** – Мильман Б.Л., Конопелько Л.А. ЖАХ, т.59, №12, с.1244-1258.

## Некоторые зарубежные публикации в области метрологии качественного анализа

- 1973** - **McLafferty** F.W. Interpretation of mass-spectra. Reading (USA).
- 1989** - **de Ruig** W.G., **Dijkstra** G., e.a. Anal. Chim. Acta. V.223, p.277-282.
- 1994** - **Ferrara** E., **Tedeschi** L., e.a. J. Anal. Toxicol. V.18, p.278.
- 1998** - **Ellison** S.L.R., **Gregory** S., **Hardcastle** W. Analyst. V.123, p.1155-1161.
- 2000** – **Hartstra** J., **Franke** J., e.a. J.Chromatogr. A. 2000. V.30, №4, P.125.
- 2002** - **Valcarse**l M., **Cardenas** S., e.a. Metrology of Qualitative Chemical Analysis. Luxembourg. 166 p.
- 2003** - **Bremser** W. e.a. Uncertainty in Semi-Qualitative Testing. Berlin.
- 2003** - **Rios** A., **Barselo** D., e.a. ACQUAL. V.8, №2, P.68

## Рабочая группа Eurachem / CITAC по метрологии качественного анализа

- **Ellison S.** (Великобритания) – председатель
- **Salit M.** (США),
- **Bremser W.** (Германия)
- **Kuselman I.** (Израиль)
- **Ferrara E.** (Италия),
- **Suchanek M.** (Чехия),
- **Pikkarainen A.L.** (Финляндия) и др.

**Принятый документ: QAWG/03/06 Eurachem/CITAC Guide:  
The Expression of Uncertainty in Qualitative Testing (September 2003)**

### **Содержание документа:**

- постановка проблемы,
- обзор методологических подходов к оценке неопределенности в качественном анализе,
- рекомендации по терминологии,
- обсуждение алгоритмов оценки достоверности идентификации,
- примеры метрологических оценок при идентификации веществ в масс-спектрометрии, в ИК-спектроскопии, в иммуноанализе.

## Предпосылки повышенного внимания к метрологии качественного анализа на рубеже XX-XXI веков

4) Возникновение в метрологии **концепции неопределенностей**, более общей, чем концепция погрешностей. В рамках этой концепции возможна оценка неопределенности идентификации с помощью вероятностных алгоритмов. Необходимо по отдельности рассчитать обе составляющие неопределенности:

$\alpha$  - вероятность ложной идентификации отсутствующего вещества

$\beta$  - вероятность необнаружения реально присутствующего вещества

*Возможны как статистические (апостериорные, эмпирические) так и априорные оценки неопределенности результатов качественного анализа.*



## Максимально допустимые уровни составляющих неопределенности для методик разного типа

Тип методики	Вероятность ложного результата	
	положительного ( $\alpha$ )	отрицательного ( $\beta$ )
Скрининг	$< 0,2$	$< 10^{-4}$
Подтверждающая	$< 10^{-4}$	$< 10^{-4}$
Референтная	$< 10^{-6}$	$< 10^{-4}$

Снижения  $\alpha$  и  $\beta$  до нужной степени можно достичь, применяя одновременно  $m$  независимых идентификационных признаков. В этом случае  $\alpha_m$  и  $\beta_m$  рассчитывают по формулам Байеса или Бернулли.

## Возможные подходы к метрологической оценке неопределенности в качественном анализе

### 1. Статистические оценки

Используют  $N$  образцов известного состава, из которых  $N_1$  не содержат  $X$ , а  $N_2$  – содержат  $X$ . С учетом критериев идентификации рассчитывают: частоту ложной идентификации  $X$  (**false positive rates, FPR, ПЛП**) и частоту необнаружения  $X$ , когда он присутствует (**false negatives rates, FNR, ЛО**).

Если в FP случаях ответ положителен и неправилен,  
в TN случаях – ответ отрицателен и правилен,  
в TP - ответ положителен и правилен,  
в FN - ответ отрицателен и неправилен:

тогда  $\alpha \approx FPR = FP / (FP+TN)$                        $\beta \approx FNR = FN / (TP + FN)$

Для 200 проб, не содержащих  $X$ , получено 8 положительных и 192 отрицательных ответа.  
Для 300 проб, содержащих  $X$ , получено 297 положительных и 3 отрицательных ответа:

$$\alpha \approx 8 / (8 + 192) = 0,04$$

$$\beta \approx 3 / (297 + 3) = 0,01$$

## Реализация статистического подхода

Считают, что FPR и FNR – случайные величины, имеющие биномиальное распределение. Тогда объем исследуемой выборки «холостых» проб (N1) при оценке  $\alpha$  по единичному признаку должен с надежностью P обеспечить появление хотя бы одной ошибки. При оценке  $\beta$  величина N2 имеет тот же порядок, что и N1.

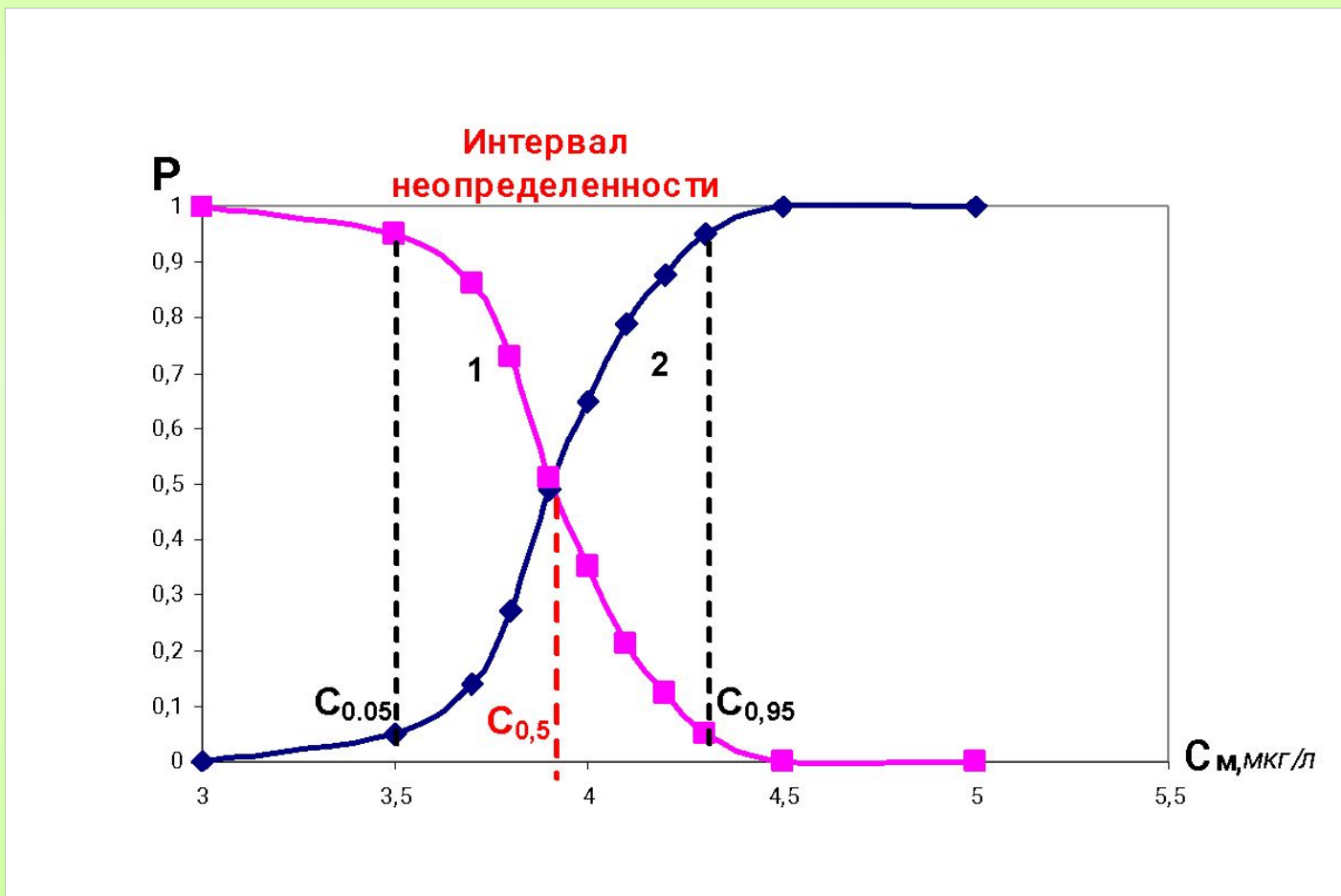
Ожидаемая вероятность ложных идентификаций	Минимальное число анализируемых проб (N1) для обеспечения заданного уровня надежности (P) при аттестации методики	
	P = 95%	P = 99%
0,005	<b>598</b>	<b>919</b>
0,01	<b>299</b>	<b>459</b>
0,05	<b>59</b>	<b>90</b>

Отсутствие ложных идентификаций при испытании 59 «холостых» проб доказывает, что  $\alpha < 0,05$  (P = 0,95).

**Чем надежнее методика идентификации - тем труднее оценить ее неопределенность!**

# Связь неопределенности идентификации с концентрацией аналита М

1 - вероятность необнаружения М ( $\beta$ ), 2 – вероятность обнаружения М.



Обнаружение тяжелых металлов с применением тест-метода  
По данным Е.А.Решетняк и соавторов (2003)

## Преимущества и ограничения статистических оценок неопределенности в качественном анализе

### Алгоритмы статистической оценки неопределенности:

- ❑ универсальны, объективны, просты;
  - ❑ хорошо отработаны в клиническом анализе;
  - ❑ пригодны для характеристики бинарных тест-методов;
  - ❑ применимы для достоверного определения пределов обнаружения;
  - ❑ не требуют информации о характере распределения экспериментальных данных и о факторах, ведущих к идентификационным ошибкам.
- 
- ❖ трудоемки и длительны;
  - ❖ требуют либо наличия множества эталонов известного состава, либо наличия референтной методики качественного анализа

**Поэтому статистические оценки следует применять для метрологической аттестации методик анализа, но не для создания или оптимизации работы СКИ**

# Возможные подходы к метрологической оценке неопределенности в качественном анализе

## 2. Априорные оценки

Исследуют характер распределения экспериментальных данных, выявляют факторы, приводящие к идентификационным ошибкам, оценивают неопределенность исходных данных. Значения  $\alpha$  и  $\beta$  рассчитывают априорно – с учетом критериев идентификации

Полученные оценки позволяют:

- быстро судить о достоверности идентификации;
- оптимизировать методики анализа и подбирать критерии идентификации;
- рассчитывать пределы обнаружения компонентов;
- не требуют стандартных образцов состава или референтных методик;
- пригодны для оптимизации работы СКИ любого типа.

Однако априорные оценки:

- не универсальны (непригодны для визуальных методов);
- зависят от выбора модели, то есть в какой-то степени субъективны;
- позволяют оценить лишь нижний предел неопределенности;
- алгоритмы оценки должны создаваться отдельно для каждого метода.

## Модель для априорной оценки неопределенности в качественном хроматографическом анализе

- Концентрации всех компонентов пробы ( $X$ ) выше, чем их пределы обнаружения с данным детектором.
- Все пики на хроматограмме хорошо разрешены, положение пика определяется только характеристикой удерживания ( $t$ ).
- $t$  – нормально распределенная случайная величина с известным стандартным отклонением  $\sigma$ . Значения  $\sigma$  для всех пиков близки.
- Математическое ожидание каждого  $t$  совпадает с константой  $t_x$  в БД (нет систематических расхождений между измеренными и табличными значениями характеристик удерживания  $X$ ).
- БД включает значения  $t_x$  для всех предполагаемых компонентов пробы и не содержит совпадающих  $t_x$  для разных  $X$ .

**В рамках этой модели и ложные идентификации, и необнаружение присутствующих компонентов определяются лишь случайными сдвигами пиков на хроматограмме пробы (неопределенностью  $t$ ).**

## Алгоритм расчета составляющих неопределенности в рамках данной модели

- 1) Если случайный сдвиг пика  $X$  из «окна»  $(t_x - d, t_x + d)$  - единственная возможная причина необнаружения  $X$ , тогда

$$\beta = 1 - 2 \Phi(d/\sigma)$$

- 2) Если случайный сдвиг пика другого компонента пробы ( $Y$ ) в указанное окно – единственная возможная причина ложной идентификации  $X$ , то

$$\alpha \approx 1 - \Phi\left(\frac{\Delta_1 - d}{\sigma}\right) - \Phi\left(\frac{\Delta_2 - d}{\sigma}\right)$$

**В приведенных формулах**

$\sigma$  - стандартное отклонение случайной величины  $t$ ;

$\Phi$  – функция Лапласа;

$\Delta_1$  and  $\Delta_2$  - оценки селективности  $t_x$  в используемой БД;

$$\Delta_1 = t_x - t_{Y_1} \quad \Delta_2 = t_{Y_2} - t_x$$

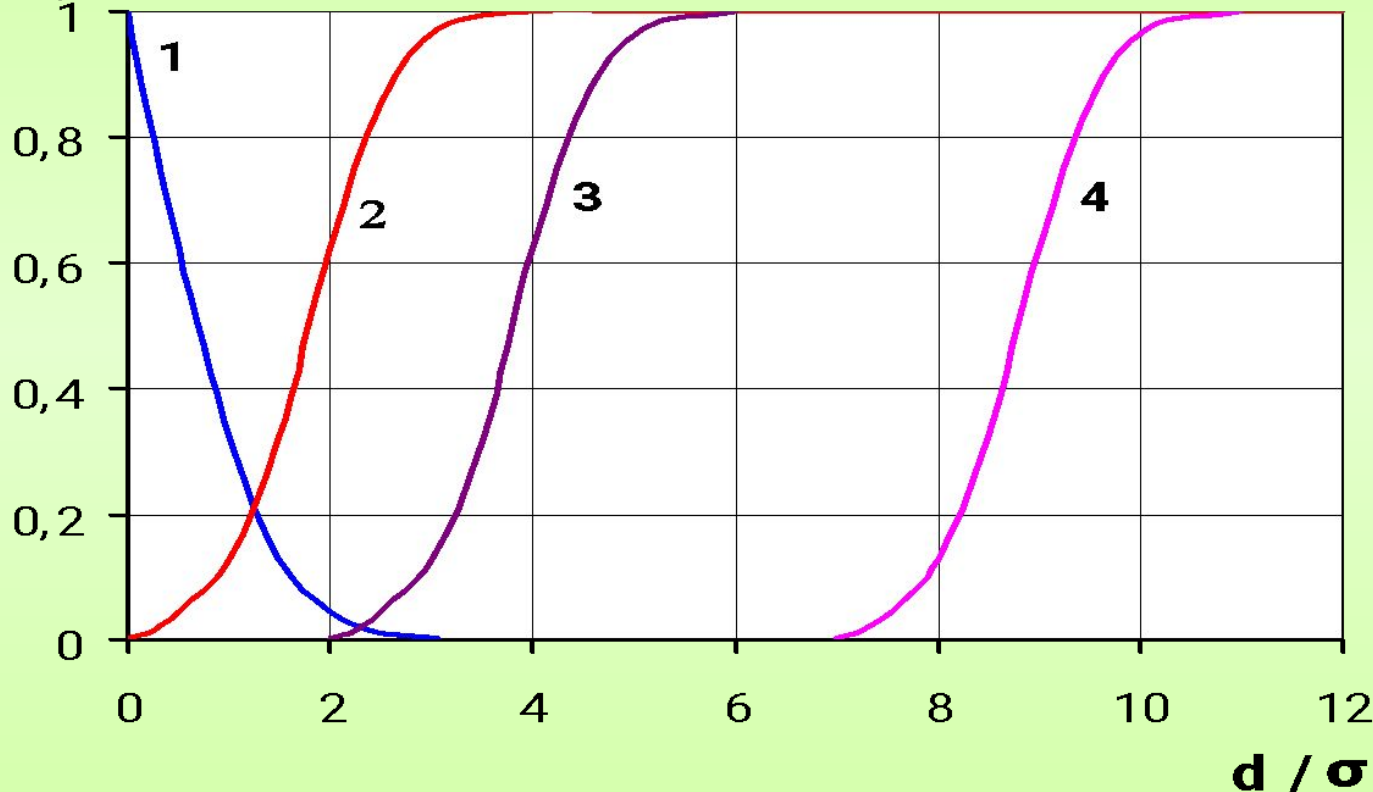
$d$  – критерий совпадения пиков.

Вещества  $Y_1$  и  $Y_2$  - ближайšie к  $X$  по характеристике  $t$ , причем  $t_{Y_2} > t_x > t_{Y_1}$ .



# 17 Оптимизация критерия $d$ при опознании веществ с разной селективностью характеристик удерживания

1 -  $\beta$ , 2,3,4 -  $\alpha$      $\Delta_1 = \Delta_2 = 3\sigma$  (2),     $\Delta_1 = \Delta_2 = 5\sigma$  (3),     $\Delta_1 = \Delta_2 = 10\sigma$  (4);  
 $\alpha, \beta$



Если при  $\sigma - \text{const}$  критерий  $d \rightarrow 0$ ,  $\beta \rightarrow 1$ ,  $\alpha \rightarrow 0$ ;  
 если же  $d \rightarrow \infty$ , тогда  $\alpha \rightarrow 1$ ,  $\beta \rightarrow 0$ . Оптимальное значение  $d = 3 \sigma$

**Число веществ, соответствующих единичному пику на хроматограмме, при разных режимах работы СКИ ( оптимизация критерия совпадения пиков )**

Пик	Время удерживания, мин.	Число опознанных веществ для разных значений критерия d				
		d = $\sigma$	d=3 $\sigma$	d=5 $\sigma$	d = 10 $\sigma$	d = 20 $\sigma$
7	13,630	1	1	1	1	1
10	15,753	0	1	1	3	3
38	36,601	0	1	2	6	12
68	47,314	1	1	4	9	19

**Однозначность идентификации достигается при d  $\approx$  3  $\sigma$**

# Априорная оценка неопределенности идентификации индивидуальных углеводородов при хроматографическом анализе бензина с помощью СКИ

Пик	t	I	I <sub>x</sub>	Name	$\alpha$	$\beta$	%
37	33,975	741,1	741,3	3,3 - dimethylhexane	0,61	0,22	0,109
38	35,718	755,3	755,3	toluene	0,02	0,21	11,066
39	35,944	756,4		unidentified			0,018
40	36,367	760,6	760,2	2,3 - dimethylhexane	0,14	0,21	0,319
41	36,541	762,0	762,1	3-ethyl -2 -methylpentane	0,03	0,24	0,078

$\alpha$  - вероятность случайной ложной идентификации данного вещества (в рамках выбранной модели);

$\beta$  - вероятность случайного необнаружения данного вещества при повторном проведении анализа той же пробы по той же методике,

# Алгоритм обратного поиска при работе систем компьютерной идентификации (СКИ)

1. **Ввод данных** ( спектра или хроматограммы пробы );
2. **Сравнение по единичному признаку** :
3. признак образца сравнения (вещества  $X$  из БД) сопоставляется с признаками пробы (с учетом критерия  $d$ );
4. **Регистрация и подсчет совпадений**. Повторение операции 2 для всех  $N$  признаков  $X$ , обнаружение  $n$  совпадений;
5. **Расчет сигнала присутствия  $X$  в пробе** (в простейшем случае сигнал  $S=n$ );
6. **Повторение операций 2-4** для всех  $M$  предполагаемых компонентов пробы;
7. **Сравнение сигналов с заранее выбранным критерием  $K$** . При  $S > K$  вещество  $X$  считается опознанным, при  $S \leq K$  отбраковывается;
8. **Формирование списка** опознанных веществ, ранжированного по  $S$ ;
9. **Вычисление количественного содержания опознанных компонентов пробы** (лишь в некоторых СКИ).
10. **Выдача результатов** пользователю

## Дополнительные операции, возможные для СКИ, в которых используются алгоритмы априорной оценки неопределенности

1. Автоматическое вычисление значений критериев - отдельно для каждого предполагаемого компонента пробы, с учетом допустимой вероятности случайных ошибок идентификации
2. Оценка достоверности идентификации каждого опознанного компонента пробы и сообщение этой информации пользователю СКИ (в виде  $\alpha$  и  $\beta$  по отдельности, либо их суммы).

## Оценка неопределенности для методик анализа, включающих $n$ единичных испытаний (в разных условиях)

$n$	$\alpha_n$	$\beta_n$	$P_\Sigma = \alpha_n + \beta_n$
1	0,2	0,20	0,40
2	0,04	0,49	0,53
5	<0,001	0,67	0,67
10	<0,001	0,89	0,89

Предполагается, что единичные вероятности  $\alpha$  и  $\beta$  постоянны.

Условие идентификации – совпадение признаков пробы и эталона в каждом испытании. Искомые вероятности определяются по формулам Бернулли:

$$\alpha_n = \alpha^n, \quad \beta_n = 1 - (1-\beta)^n$$

Если  $n \rightarrow \infty$ , то  $\alpha_n \rightarrow 0$ , но при этом  $\beta_n \rightarrow 1$

В ходе многократных испытаний суммарная неопределенность **возрастает!**

## Оценка $\alpha_n$ для методов, основанных на подсчете количества спектральных совпадений пробы и эталона

Модель предполагает равную точность измерения длин волн в интервале  $(\lambda_1, \lambda_2)$ , где спектр пробы содержит  $M$  линий, а спектр эталона  $X$  -  $N$  линий.

Критерий совпадения  $d = \Delta\lambda$  одинаков для всех линий. Условие идентификации – реальное число совпадений больше заданного критерия  $K$ .

Вероятность случайного совпадения одной линии в спектре пробы и какой-либо из линий эталонного спектра:

$$\alpha = \frac{2 \cdot d \cdot N}{|\lambda_2 - \lambda_1|}$$

Вероятность одновременного случайного совпадения  $n$  линий в спектре пробы с линиями эталонного спектра вычисляется по формулам Бернулли:

$$P_{M,n} = \frac{M! \cdot \alpha^n (1 - \alpha)^{M-n}}{n! (M - n)!}$$

Вероятность случайного совпадения спектров пробы и эталона более, чем по  $K$  линиям :

$$\alpha_n = 1 - \sum_0^K P_{M,n}$$

$\alpha_n$  - искомая оценка вероятности случайной ложной идентификации для любого целочисленного  $K$ .

# Расчет критерия идентификации в спектральном анализе

24

Для снижения  $\alpha_n$  до желаемого уровня рекомендуется критерий:

$$n_{\text{крит}} \approx M \alpha + t_{\text{кр}} \sqrt{M \cdot \alpha \cdot (1 - \alpha)}$$

где  $t_{\text{кр}}$  находят из условия  $\Gamma(t) = 1 - \alpha_n$ , где  $\Gamma(t)$  - интеграл Гаусса. Так, для  $\alpha_n = 0,05$   $t_{\text{кр}} = 1,65$ .

Пример: в спектре пробы 40, а в спектре эталона X - 20 линий. Оба спектра сняты в интервале шириной  $|\lambda_2 - \lambda_1|$ . Округленные значения критерия  $n_{\text{кр}}$ , обеспечивающие  $\alpha_n < 0,05$ , при разной точности измерения длин волн равны:

$\Delta\lambda$ , нм	$\alpha$	$n_{\text{кр}}$
0,01	0,002	1
0,05	0,010	2
0,20	0,040	4

При постоянной величине  $\Delta\lambda$  значения  $n_{\text{кр}}$  для разных веществ должны быть различны, чтобы обеспечить одну и ту же заданную надежность идентификации.

Причина – разное число линий в эталонных спектрах разных веществ (разные N).



## Результаты компьютерного качественного анализа бинарной смеси ПАУ

Расшифровка спектра низкотемпературной люминесценции пробы.

<b>Компонент X</b>	<b>N</b>	<b>n</b>	<b>K = n<sub>крит</sub></b>
<u><b>Фенантрен</b></u>	22	14	4
<u><b>Трифенилен</b></u>	26	25	5
Хризен (?)	14	4	3

**N** - число линий в эталонном спектре X;

**n** - число линий X, обнаруженных ( $d = 0,1$  нм) в спектре пробы,

**K** (или  $n_{\text{крит}}$ ) – оптимизированный критерий идентификации,

рассчитанный для  $\alpha_n = 0,05$ ;

Названия действительно присутствующих ( $10^{-7}$  г/мл) компонентов подчеркнуты.

197 предполагаемых компонентов, для которых наблюдались совпадения линий, но у которых  $n < K$  (i.e.  $\alpha_n > 0,05$ ), - отбракованы.



## Некоторые нерешенные проблемы метрологии качественного анализа

### **В рамках статистического подхода:**

унификация терминологии и вычислительных алгоритмов,  
разработка специальных стандартных образцов,  
создание алгоритмов для оценки пределов обнаружения с учетом  $\alpha$  ,  
создание алгоритмов для оптимизации критериев идентификации.

### **В рамках априорного подхода:**

создание обоснованных теоретических моделей,  
исследование бюджета неопределенности,  
учет селективности поисковых признаков

**Наиболее перспективным направлением исследований представляется постепенное сближение обоих подходов**

- 1) *Вершинин В.И., Топчий В.А., Наумов С.Е.* Число спектральных совпадений как критерий идентификации компонентов пробы. Ж.аналит.химии. 1987. Т.42, № 5, с.837.
- 2) *Вершинин В.И.* Методология компьютерной идентификации веществ с применением информационно-поисковых систем. Ж.аналит.химии. 2000, Т.55, № 5, С.468.
- 3) *Соколова О.В., Ильичева Н.Б., Вершинин В.И.* Достоверность компьютерной идентификации углеводов при хроматографическом анализе бензинов. Аналитика и контроль. 2000, № 4, с.363.
- 4) *Вершинин В.И., Топчий В.А., Медведовская И.И.* Критерии совпадения пиков в качественном хроматографическом анализе. Ж.аналит.химии. 2001, Т.56, № 4, с.367.
- 5) *Вершинин В.И., Дерендяев Б.Г., Лебедев К.С.* Компьютерная идентификация органических соединений. 2002. М. “Академкнига”, 197 с.
- 6) *Медведовская И.И., Вершинин В.И.* Сходимость характеристик удерживания как условие достоверной идентификации. Зав.лаборатория. 2004. Т.70, № 7, с.18.
- 7) *Vershinin V.I.* A priori method of evaluating uncertainties in qualitative chromatographic analysis:(probabilistic approach). **Accreditation and Quality Assurance**. 2004. V.9, № 7.P.415 – 418.