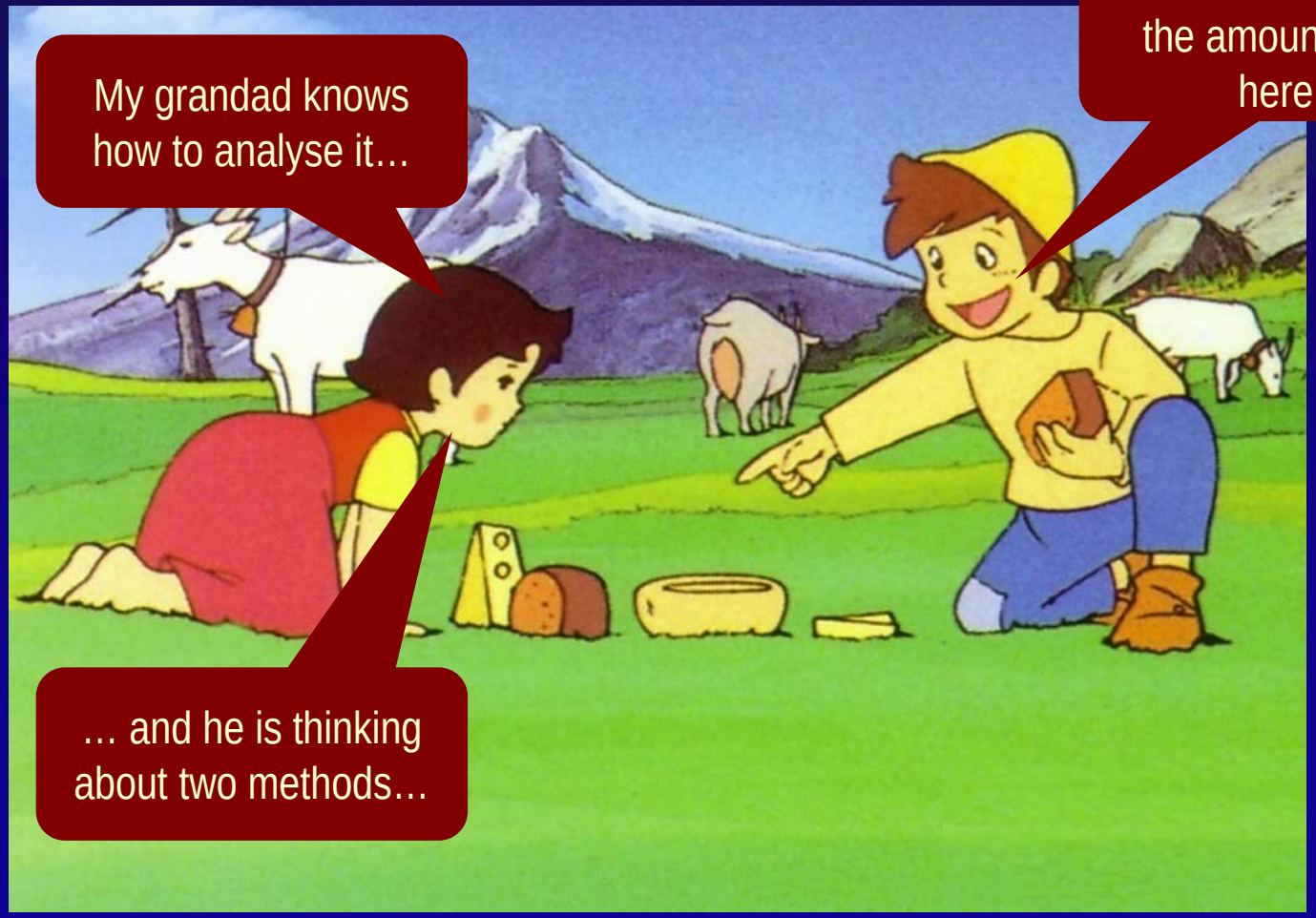
Data analysis in chromatography

The old battle of Bayesian vs. frequentist revisited

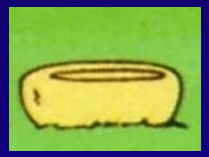
Gabriel Vivó Truyols
Van ‘t Hoff institute for molecular sciences
University of Amsterdam

Statistical inference

How much iron is in my milk?



Statistical inference

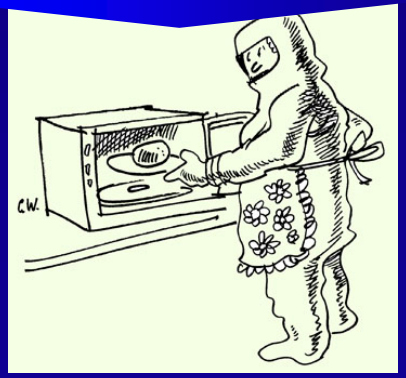


Iron?

Reference method: acid reaction

Costly (we know that works)

New method: microwave degradation



Faster (but… does it work?)

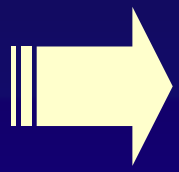
Is the new method providing the same results as the reference method?

Statistical inference

Introduction

Is the new method working correctly? ????

Go to the lab, perform an experiment on a sample of n objects…



… conclusion: "the new method provides the same results as the reference" or "the new method does not work"

Limited information (information only about a sample)

Statistical inference

General conclusion (information about the population)

Statistical inference using frequentist approach

Frequentist approach

Is the new method working correctly?

Go to the lab, perform an experiment on a sample of n objects…

… conclusion: "the new method provides the same results as the reference" or "the new method does not work"

The iron concentration of the pill is for sure 100 mg.
We measure the pill 6 times with the new method. The results are:
98.9; 100.3; 99.7; 99.0; 100.6; 98.6

The mean value is: 99.5

Statistical inference

Is the true mean 100?

… "true mean" means here "population mean", μ, or the mean obtained with infinite replications with the new method

Statistical inference using frequentist approach

Frequentist approach

Is the new method working correctly?

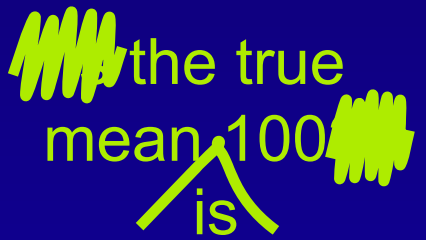
Go to the lab, perform an experiment on a sample of n objects…

… conclusion: "the new method provides the same results as the reference" or "the new method does not work"

The iron concentration of the pill is for sure 100 mg.
We measure the pill 6 times with the new method. The results are:
98.9; 100.3; 99.7; 99.0; 100.6; 98.6

The mean value is: 99.5

Suppose that the true mean is 100… what is the chance of obtaining this data (or more extreme)?



Statistical inference using frequentist approach

Frequentist approach

Is the new method working correctly?

Go to the lab, perform an experiment on a sample of n objects…

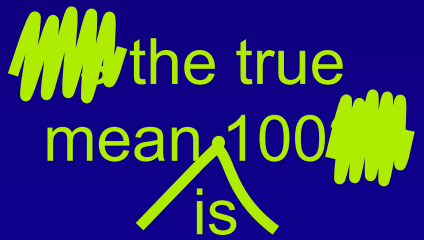
… conclusion: "the new method provides the same results as the reference" or "the new method does not work"

The irr… …ill is

We m… … the ne… …:
98.9; …98.6

The probability (p-value) of obtaining this result (or more extreme) is 21%

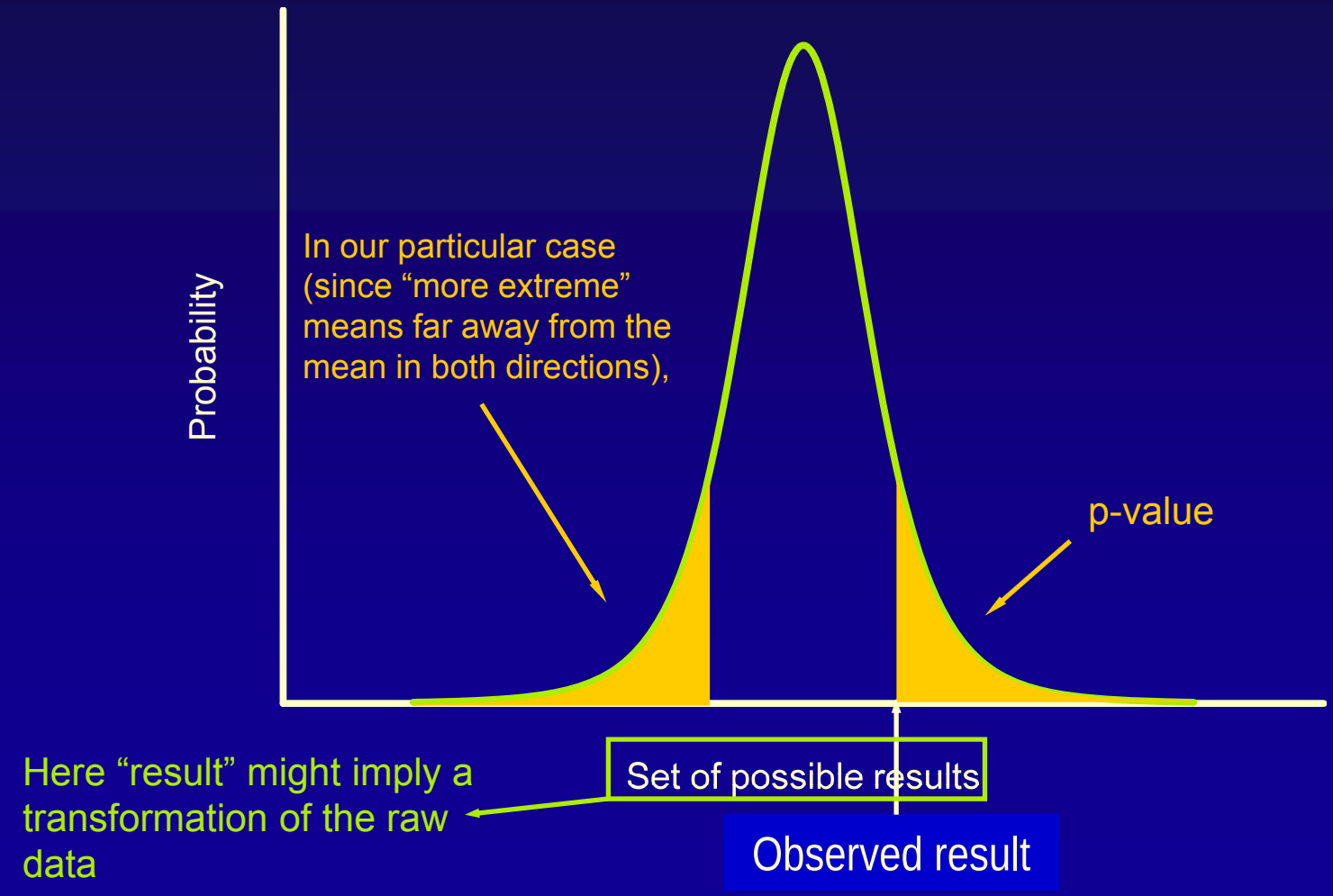
Suppose that the true mean is 100… what is the chance of obtaining this data (or more extreme)?



As this probability (p-value) is not low enough (below α), we don't have enough "proof" to reject the fact that the true mean is 100

Statistical inference using frequentist approach

Frequentist approach



Statistical inference using frequentist approach

Frequentist approach

p-value is… *… the probability of obtaining the data obtained or more extreme supposing the null hypothesis true*

p-value is NOT… *… the probability that the null hypothesis is true, given the data.*

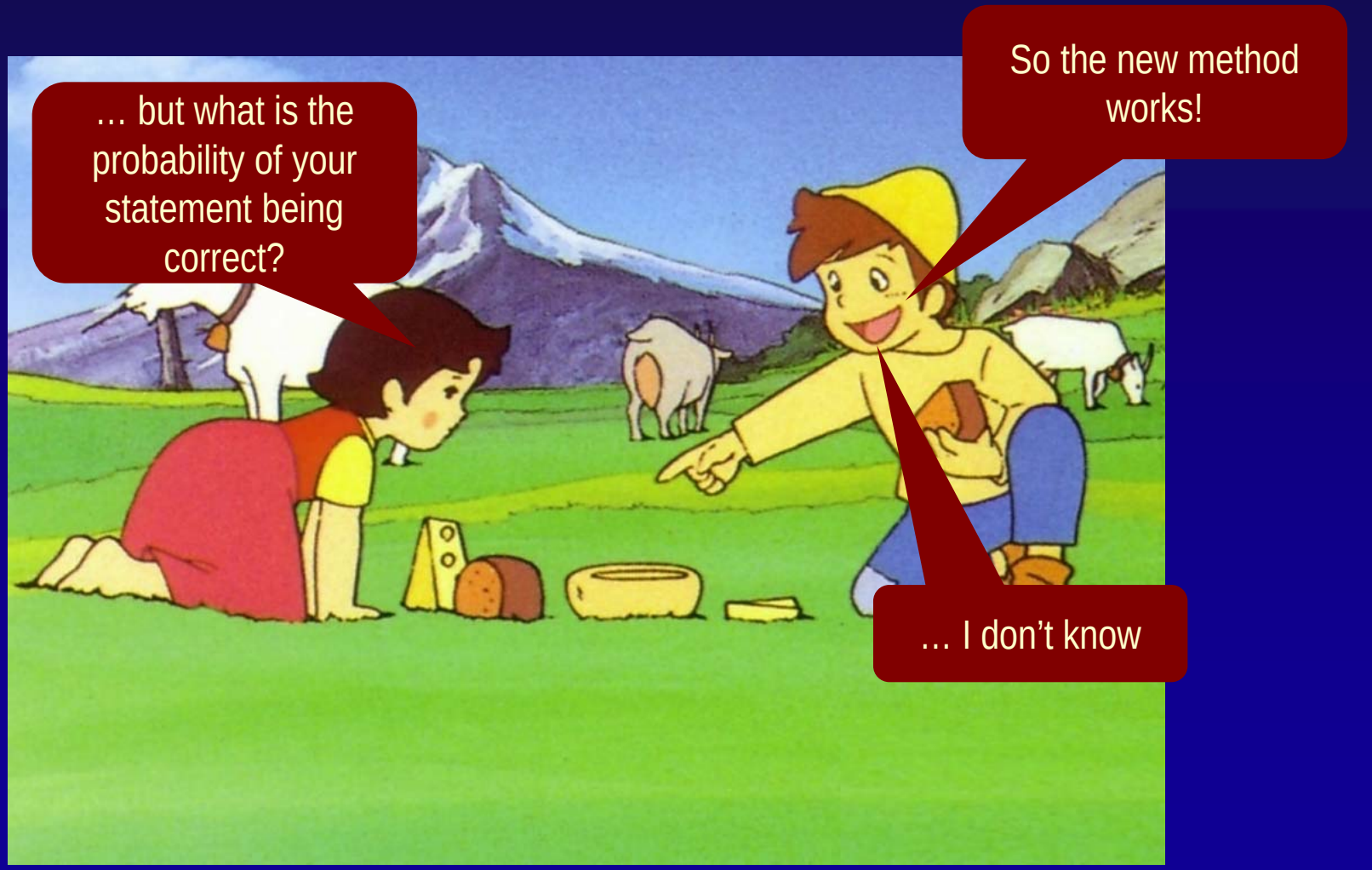
So, a p-value does NOT inform us AT ALL about the validity of a certain hypothesis…

However

… by following the procedure of rejecting the null hypothesis when p-value<α the probabilities of type-I and type-II error can be calculated.

Statistical inference

Is the new method working?

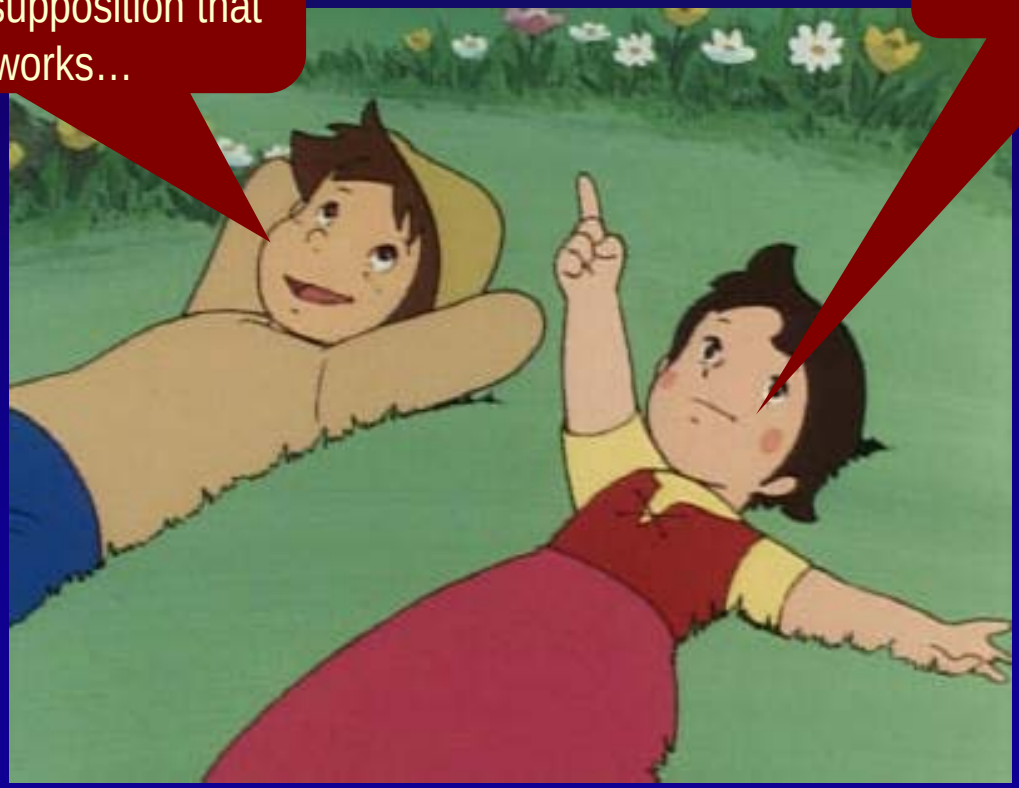


Statistical inference

Are there alternatives?

… I can only tell you the probability of obtaining this data or more extreme under the supposition that the method works…

I think there are alternatives…



The Bayes' rule

Bayesian approach

General form

$$p(H_i|\mathrm{D}) = p(\mathrm{D}|Hi) \times \frac{p(H_i)}{p(D)}$$

- $p(H_i|\mathrm{D})$: Posterior probability for a certain hypothesis (H_i)
- $p(\mathrm{D}|Hi)$: Likelihood (probability of the data given the hypothesis); Data
- $\frac{p(H_i)}{p(D)}$: Normalization constant; Prior probability for a certain hypothesis (H_i)

Odds form (only two hypotheses considered, H_0 and H_1)

$$\frac{p(H_0|\mathrm{D})}{p(H_1|\mathrm{D})} = \frac{p(\mathrm{D}|H_0)}{p(\mathrm{D}|H_1)} \frac{p(H_0)}{p(H_1)}$$

- $\frac{p(H_0|\mathrm{D})}{p(H_1|\mathrm{D})}$: Posterior odds
- $\frac{p(\mathrm{D}|H_0)}{p(\mathrm{D}|H_1)}$: Likelihood ratio (sometimes called Bayes' factor)
- $\frac{p(H_0)}{p(H_1)}$: Prior odds

The Bayes' rule

Bayesian approach

Let's consider the two competing hypothesis, and let's see how our "prior odds" about the validity of the two hypothesis is updated when we take the experimental data (D) into consideration

The method works H_0

The method does not work H_1

Bayes' rule

$$\frac{p(H_0|\mathrm{D})}{p(H_1|\mathrm{D})} = \frac{p(\mathrm{D}|H_0)}{p(\mathrm{D}|H_1)} \frac{p(H_0)}{p(H_1)}$$

Posterior odds

Likelihood ratio (sometimes called Bayes' factor)

Prior odds

The Bayes' rule. The prior odds.

Bayesian approach

The method works H_0

The method does not work H_1

$p(H_0)$ → "Prior" probability that the method works

$p(H_1)$ → "Prior" probability that the method doesn't work

Before inspecting the data,
what is my "preference"
for each of the
two competing hypothesis?

???

If I don't have any "preference", $p(H_0) = p(H_1) = 0.5$, and the prior odds is 0.5/0.5=1

The Bayes' rule. The likelihood ratio.

Bayesian approach

The method works H_0

The method does not work H_1

$p(\mathrm{D}|H_0)$ → Probability of obtaining the data under the supposition that H_0 is true

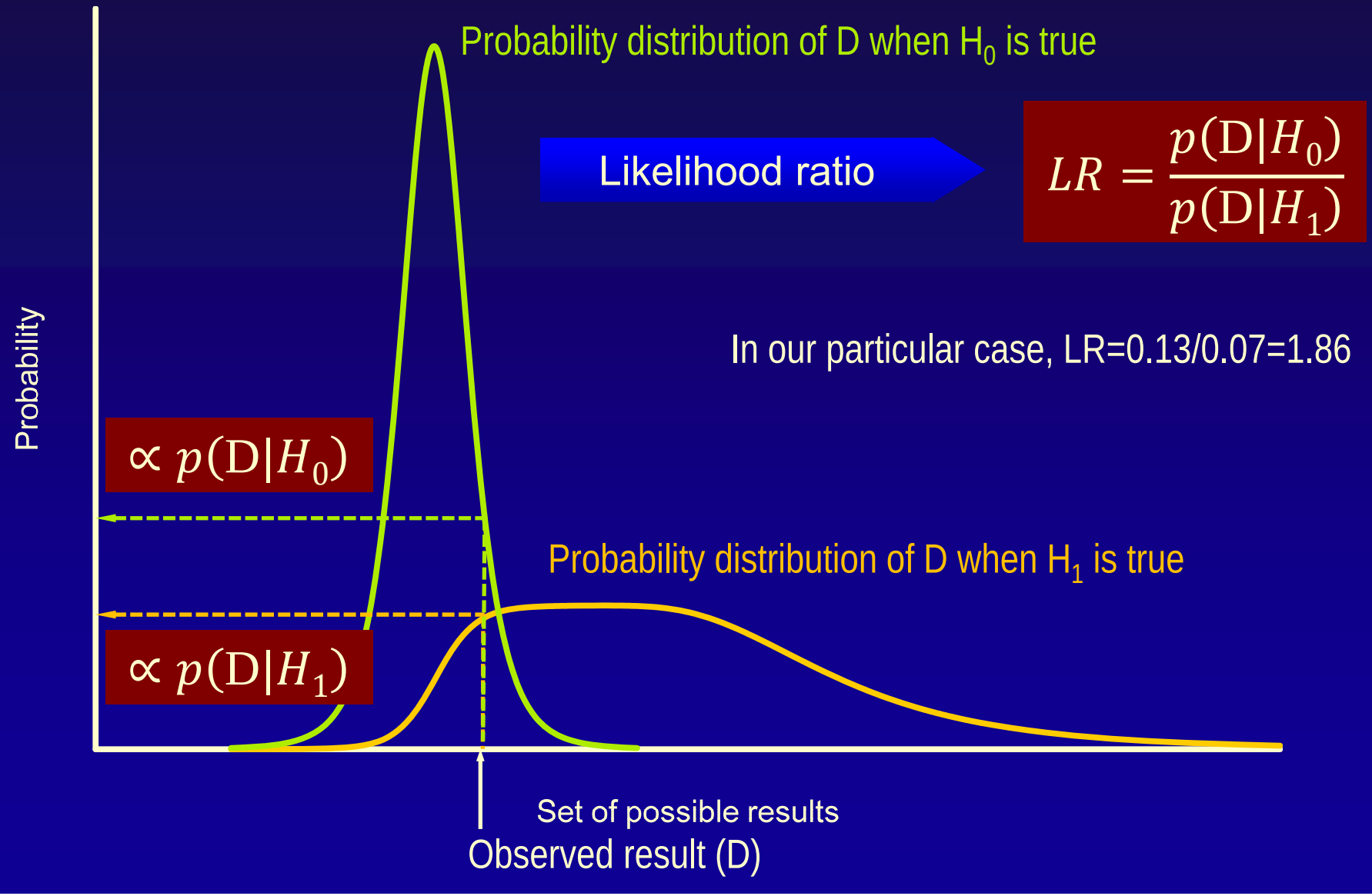
$p(\mathrm{D}|H_1)$ → Probability of obtaining the data under the supposition that H_1 is true

Data (D)

We measure the pill 6 times with the new method. The result is:
98.9; 100.3; 99.7; 99.0; 100.6; 98.6

The Bayes' rule. The likelihood ratio.

Bayesian approach



The Bayes' rule. The posterior odds.

Bayesian approach

Posterior odds		Likelihood ratio		Prior odds
$\frac{p(H_0\|\mathrm{D})}{p(H_1\|\mathrm{D})}$	=	$\frac{p(\mathrm{D}\|H_0)}{p(\mathrm{D}\|H_1)}$	x	$\frac{p(H_0)}{p(H_1)}$
1.86	=	1.86	x	1

Once I consider the data, what is my "preference" for each of the two competing hypothesis?

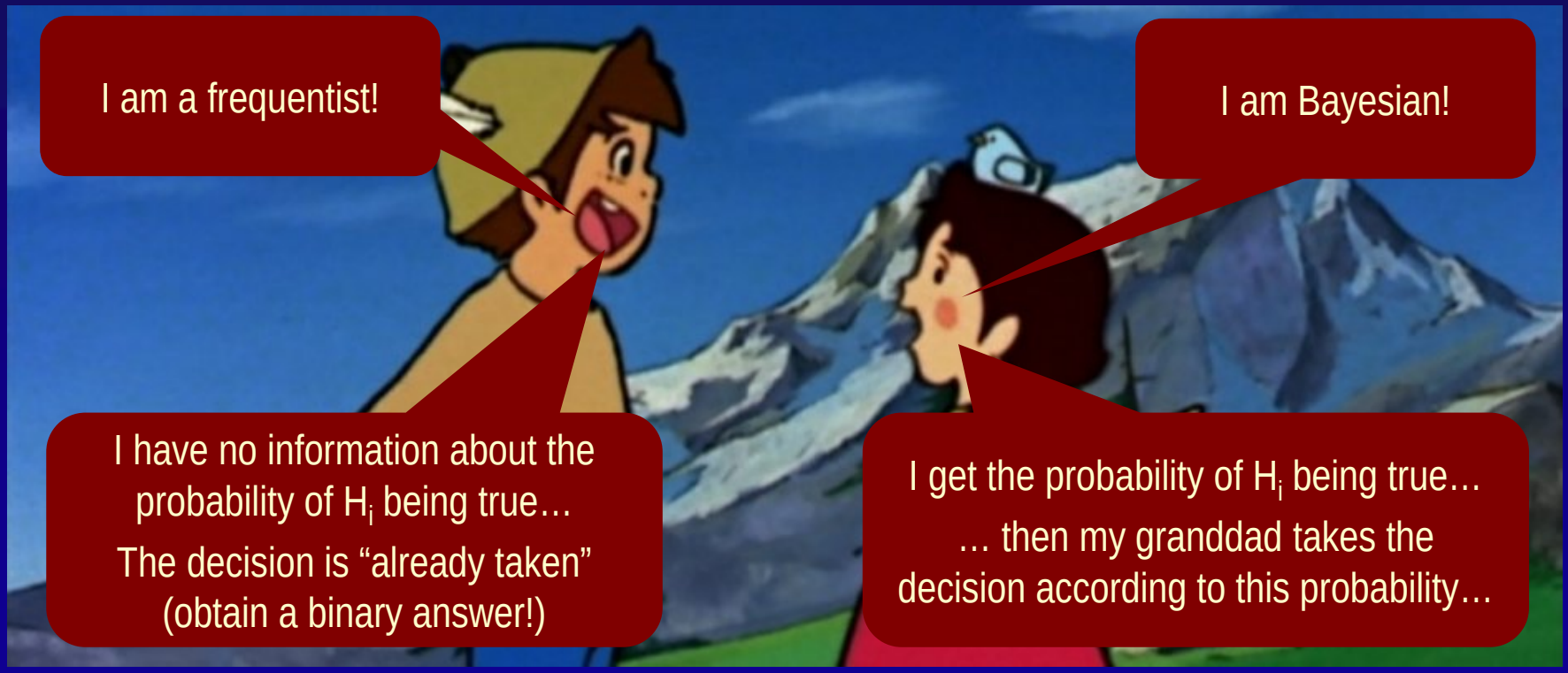
Now the probability of H_0 is 1.86 times higher than the probability of H_1

Before inspecting the data, what is my "preference" for each of the two competing hypothesis?

No preference (prior odds = 1)

Frequentist vs. Bayesian

Lesson 1



Frequentist vs. Bayesian

Incorporating new evidence…

… I am wondering how I could deal with the incorporation of new experiments (is my decision going to change?)

I think I have an idea!

The Bayes' rule. Incorporating new evidence…

Bayesian approach

Evidence D_1 *… my first set of experiments*

Posterior odds | Likelihood ratio | Prior odds

$$\frac{p(H_0|D_1)}{p(H_1|D_1)} = \frac{p(D_1|H_0)}{p(D_1|H_1)} \times \frac{p(H_0)}{p(H_1)}$$

Evidence D_2 *… a second set of experiments (or other information)*

Posterior odds | Likelihood ratio | Prior odds

$$\frac{p(H_0|D_1, D_2)}{p(H_1|D_1, D_2)} = \frac{p(D_2|H_0, D_1)}{p(D_2|H_0, D_1)} \times \frac{p(H_0|D_1)}{p(H_1|D_1)}$$

The Bayes' rule. Incorporating new evidence…

Bayesian approach

Posterior odds

$$\frac{p(H_0|\mathrm{D}_1,\mathrm{D}_2)}{p(H_1|\mathrm{D}_1,\mathrm{D}_2)}$$

=

Likelihood ratio

$$\frac{p(\mathrm{D}_2|H_0,\mathrm{D}_1)}{p(\mathrm{D}_2|H_0,\mathrm{D}_1)}$$

x

Prior odds

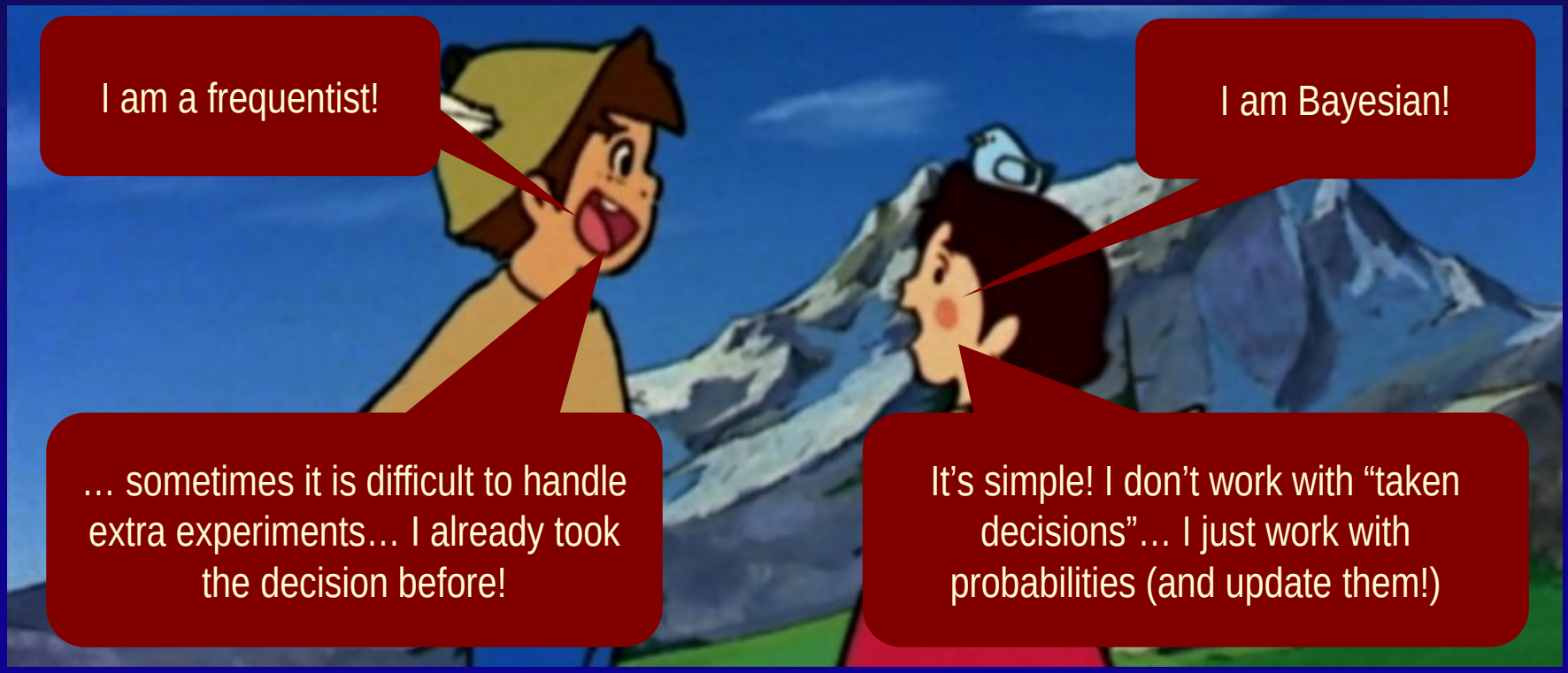
$$\frac{p(H_0|\mathrm{D}_1)}{p(H_1|\mathrm{D}_1)}$$

If D_1 and D_2 are independent…

$$\frac{p(\mathrm{D}_2|H_0)}{p(\mathrm{D}_2|H_0)}$$

Frequentist vs. Bayesian

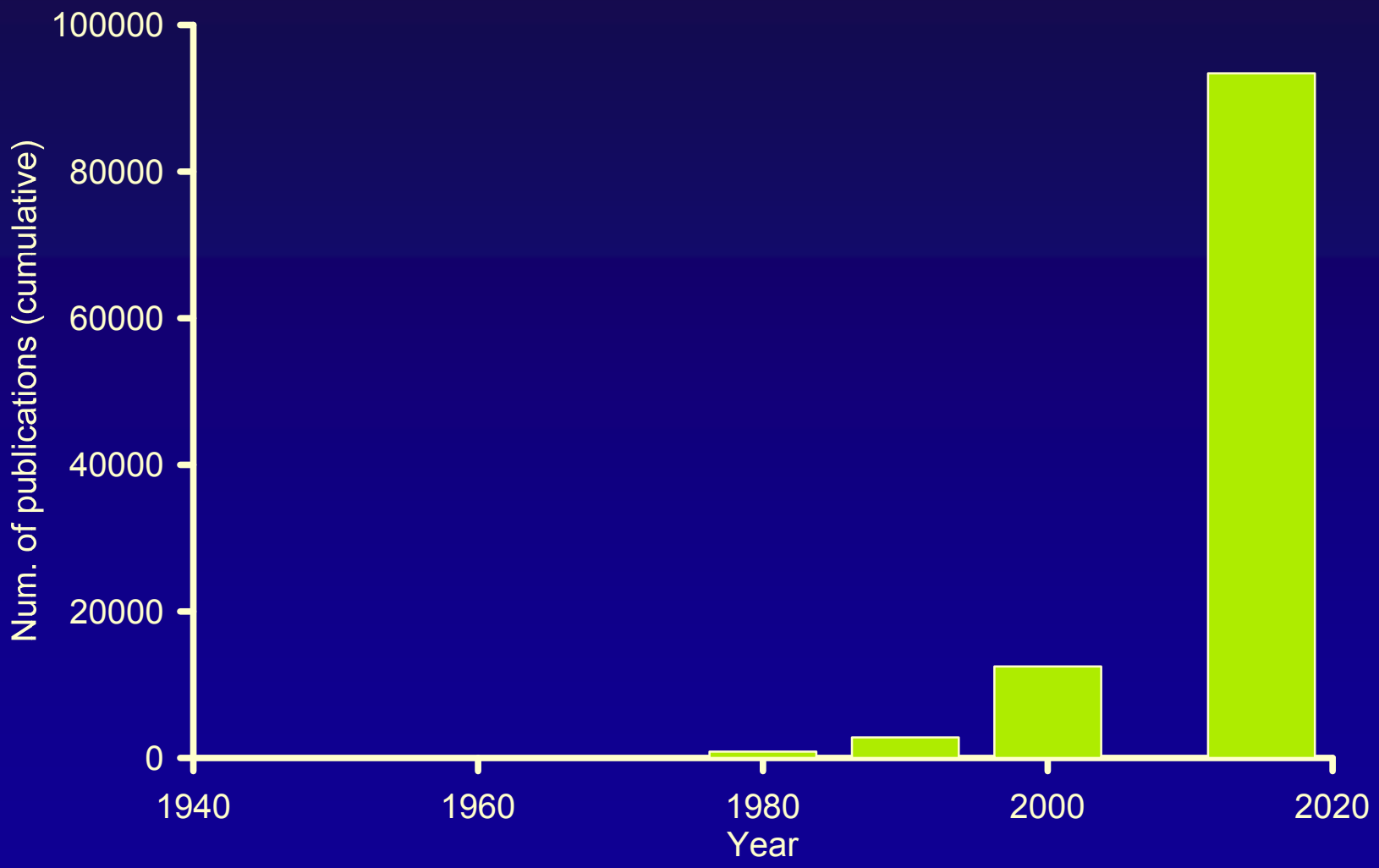
Lesson 2



Why Bayesian?

I am not the first ...

Why Bayesian?



Automation is our challenge now...

Why Bayesian?

Evolution of the instrumentation

1 bit

GC: 1 KB

HPLC: 1KB

GC-MS; ~100 KB

HPLC-DAD; ~100 KB

HPLC-MS;
1 MB

GCxGC: ~100 KB

GCxGC-MS;
LCxLC-MS;
1 GB/hour

GCxGC-HRMS;
LCxLC-HRMS;
15 GB/hour

Automation using a frequentist approach

Why Bayesian?

Data + *Algorithm* = *Information*

$$\sum \theta \sin \epsilon$$

There is a chromatographic peak at t_R=12 min.

These 10 peaks in these 10 chromatograms belong to the same compound.

… etc.

In a sense, the machines are "taking responsibility" on the decision…

… and only the final result is shown.

Automation using a frequentist approach

Why Bayesian?

Final result

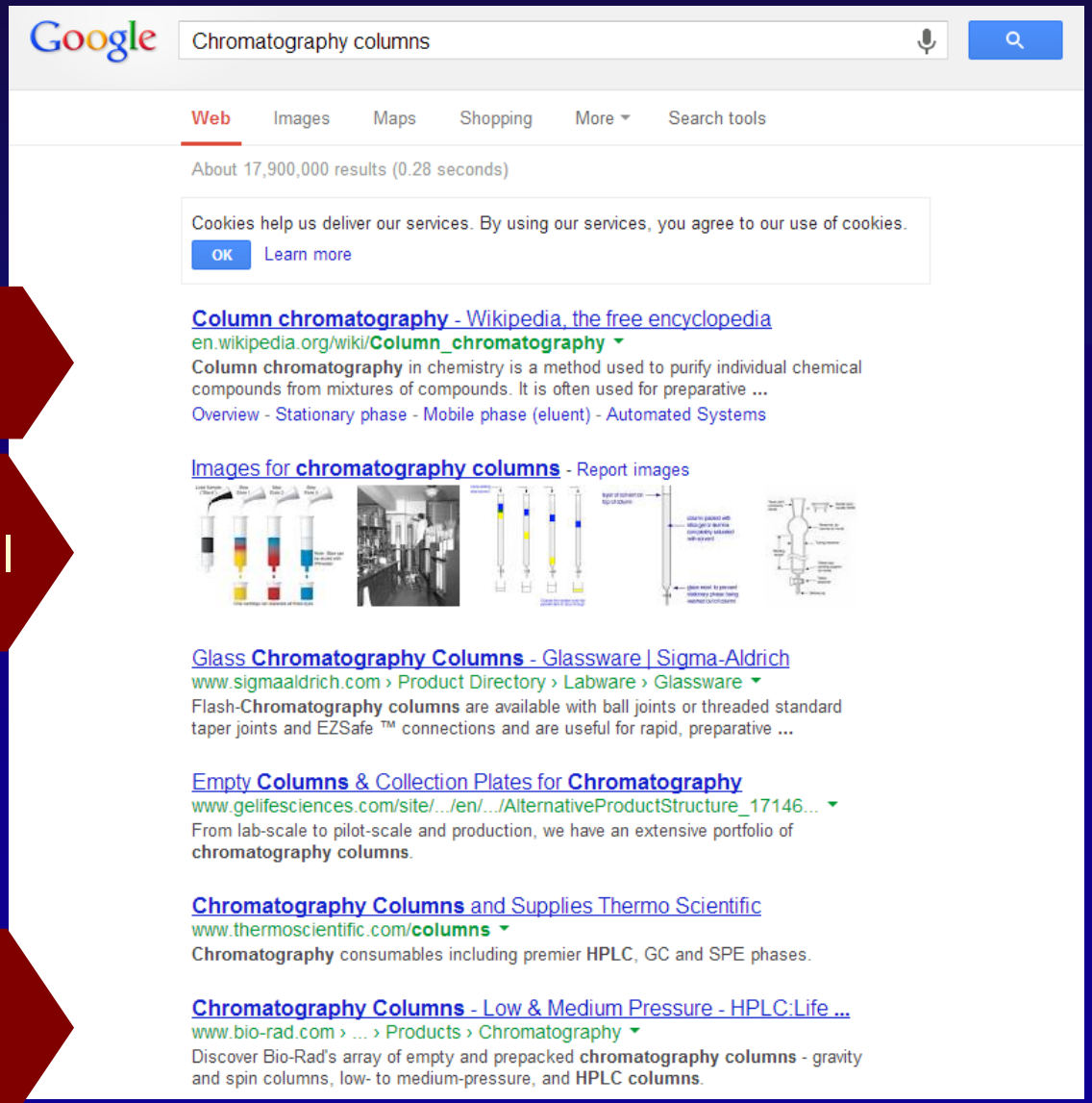
Automation using a "Bayesian" approach

Why Bayesian?

Most likely result

Less likely result (but still possible)

…



Automation using a Bayesian approach

Why Bayesian?

Data

87% chance: There is a peak centred at t_R=12 min.

13% chance: There is no peak centred at t_R=12 min.

20% chance: Peaks 1-9 belong to the same compound. Peak 10 is different.

15% chance: Peaks 1-5,7-9 belong to the same compound. Peaks 6 and 10 are different.

… etc.

Information

There is a chromatographic peak at t_R=12 min.

These 10 peaks in these 10 chromatograms belong to the same compound.

a collection of all possibilities (ranked by their probability)

In a sense, the machines are ~~"taking~~ not responsibility" on the decision…

… and only the final result is shown.

It is up to the chromatographer to take the final decision

Bayesian statistics in court

Why Bayesian?

Posterior odds = Likelihood ratio x Prior odds

$$\frac{p(H_0|\mathrm{D})}{p(H_1|\mathrm{D})} = \frac{p(\mathrm{D}|H_0)}{p(\mathrm{D}|H_1)} \times \frac{p(H_0)}{p(H_1)}$$

The suspect is innocent

The suspect is guilty

The scientist does not decide upon the validity of H_0 or H_1, only calculates the likelihood ratio (the value of the evidence)

Bayesian statistics in data automation

Why Bayesian?

Posterior odds = Likelihood ratio x Prior odds

$$\frac{p(H_0|\mathrm{D})}{p(H_1|\mathrm{D})} = \frac{p(\mathrm{D}|H_0)}{p(\mathrm{D}|H_1)} \times \frac{p(H_0)}{p(H_1)}$$

I take the decision!!

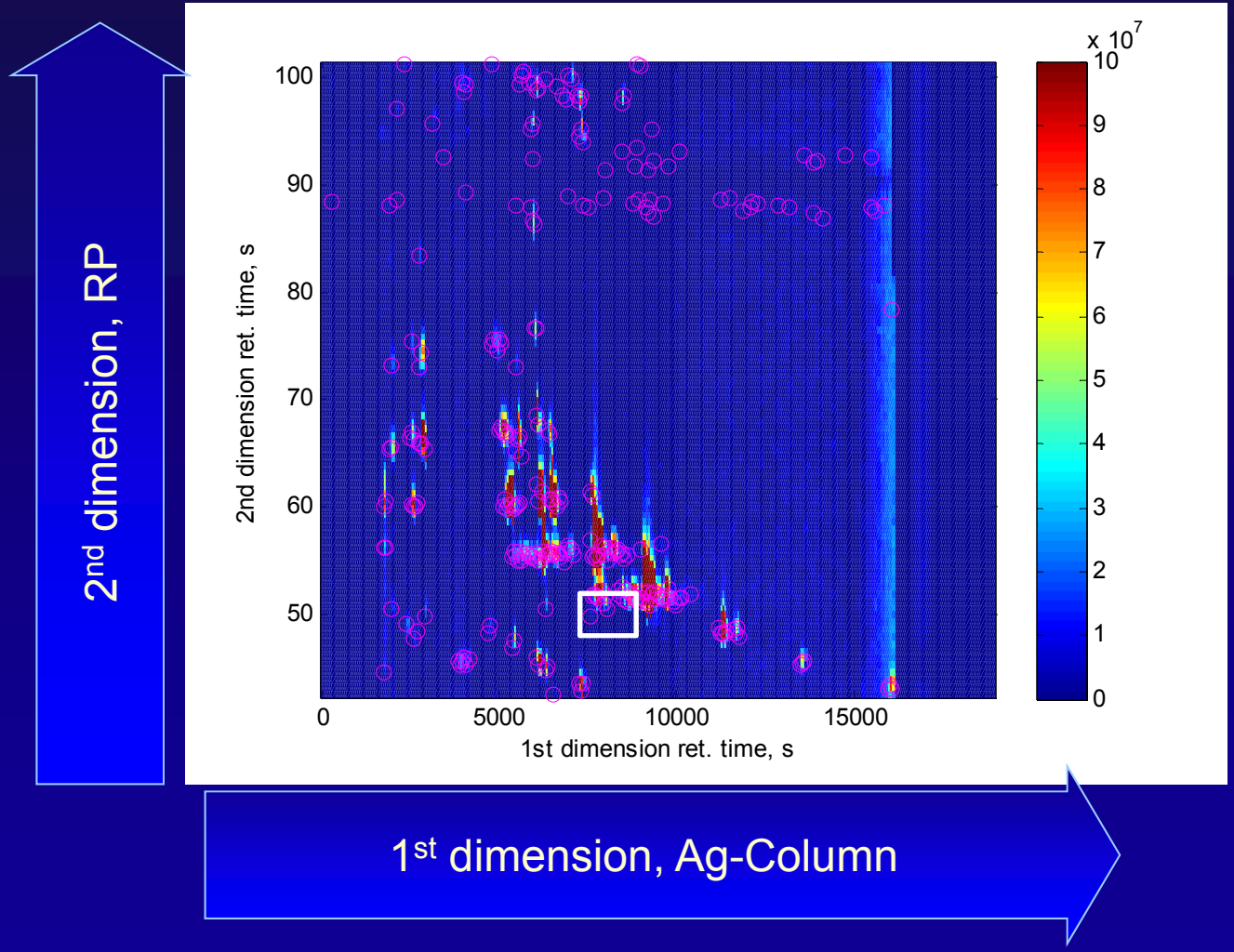
The data is only used to "update" our prior probability on a situation (**but the decision is not "taken" by the algorithm**)

Prior experiments
Bibliographic information

Some practical applications in chromatography

Example I: peak detection in two-dimensional chromatography

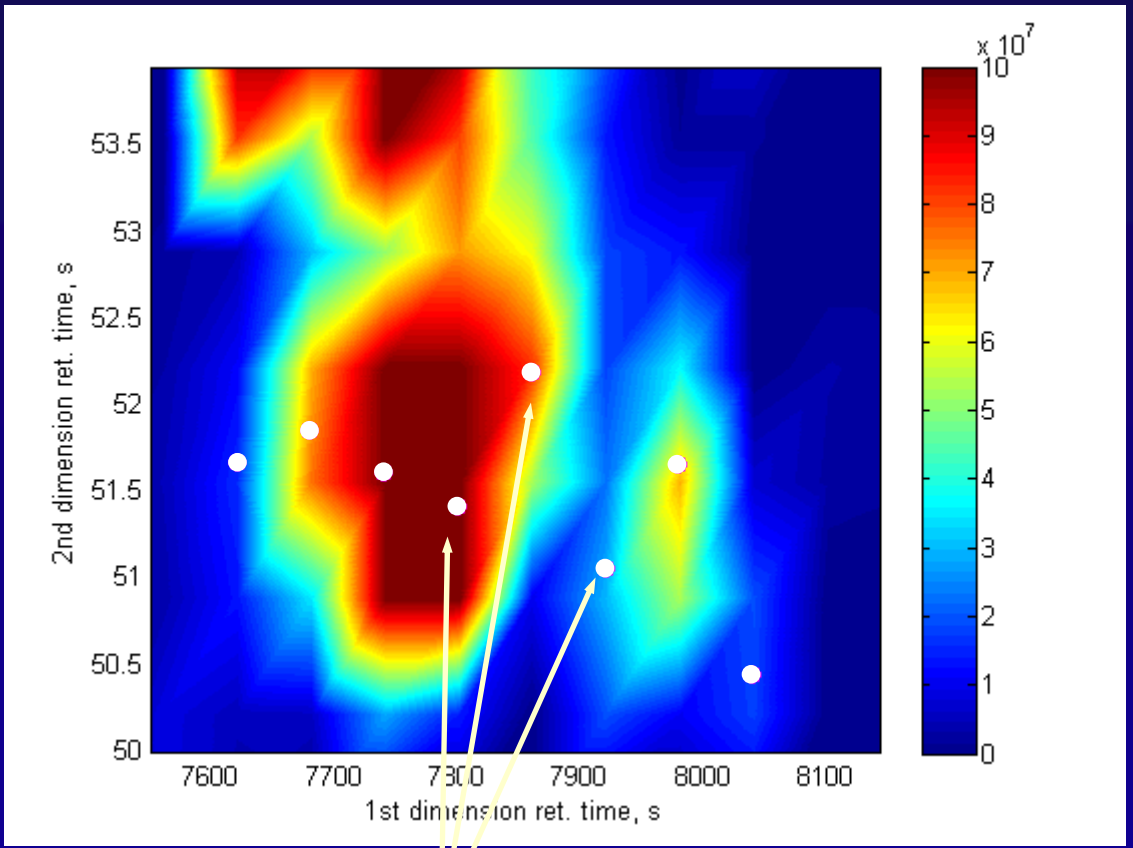
Example I



Data courtesy of Teris van Beek, University of Wageningen (NL)

Example I: peak detection in two-dimensional chromatography

Example I

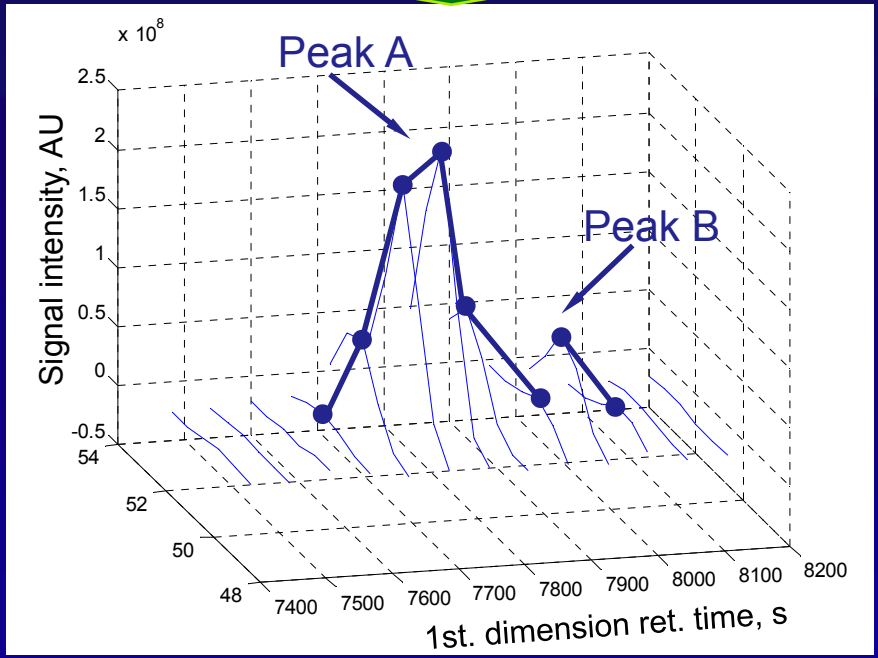


Each of these dots corresponds to a detected peak

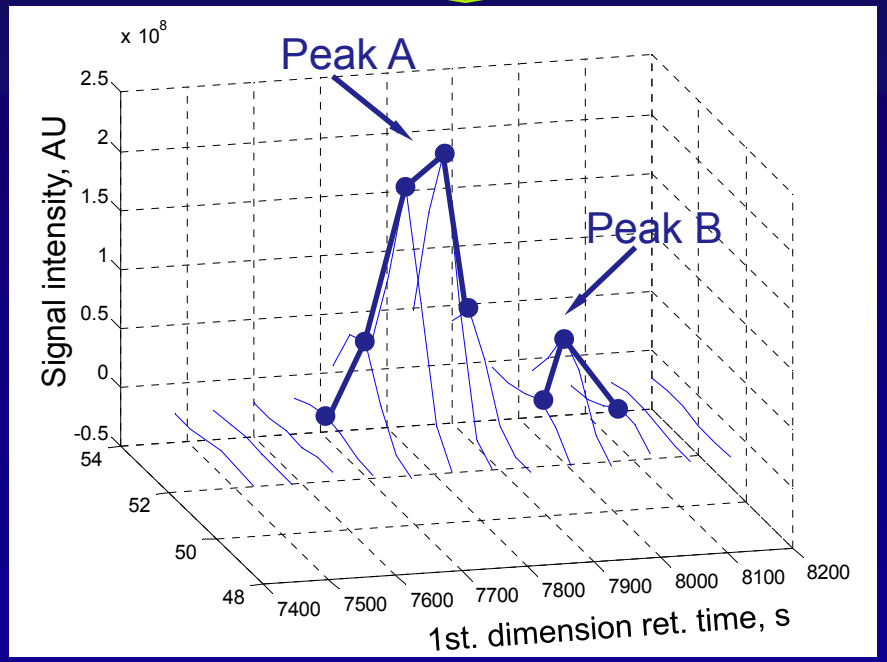
Example I: peak detection in two-dimensional chromatography

Example I

Possibility 1

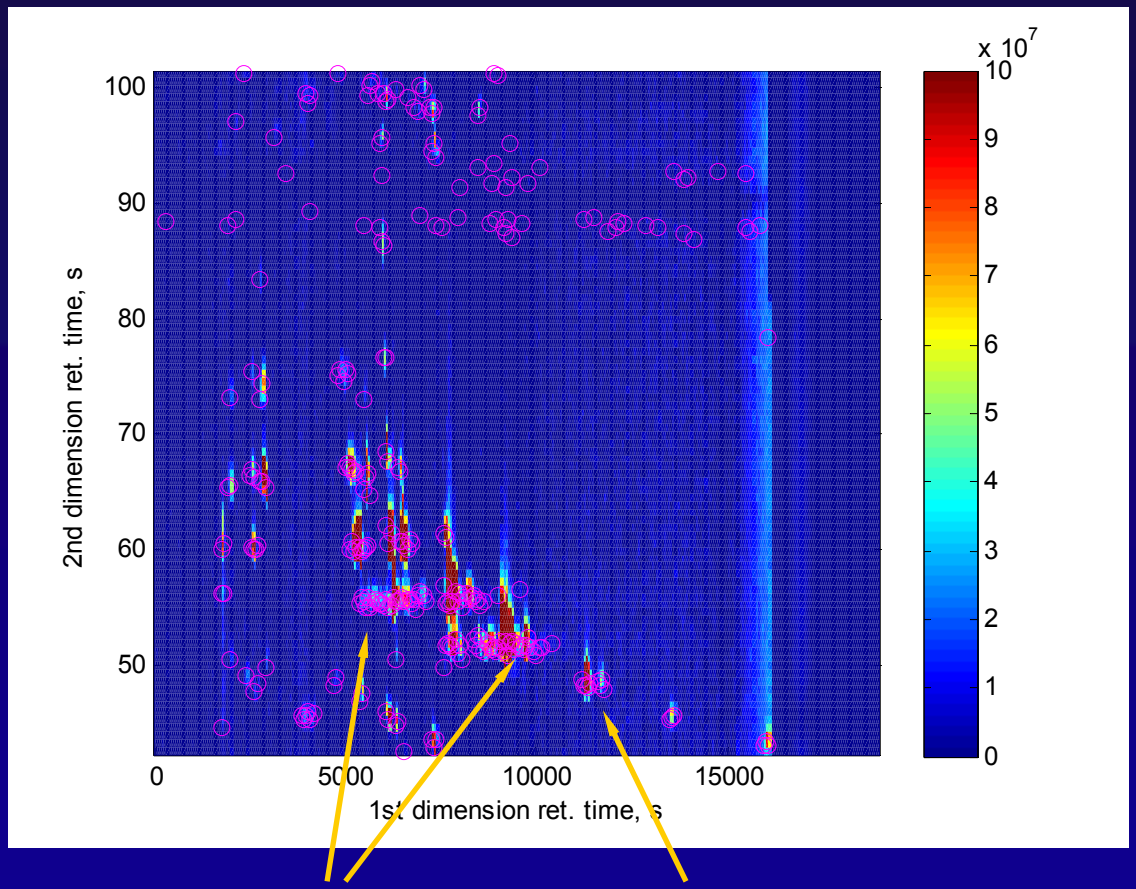


Possibility 2



Example I: peak detection in two-dimensional chromatography

Example I



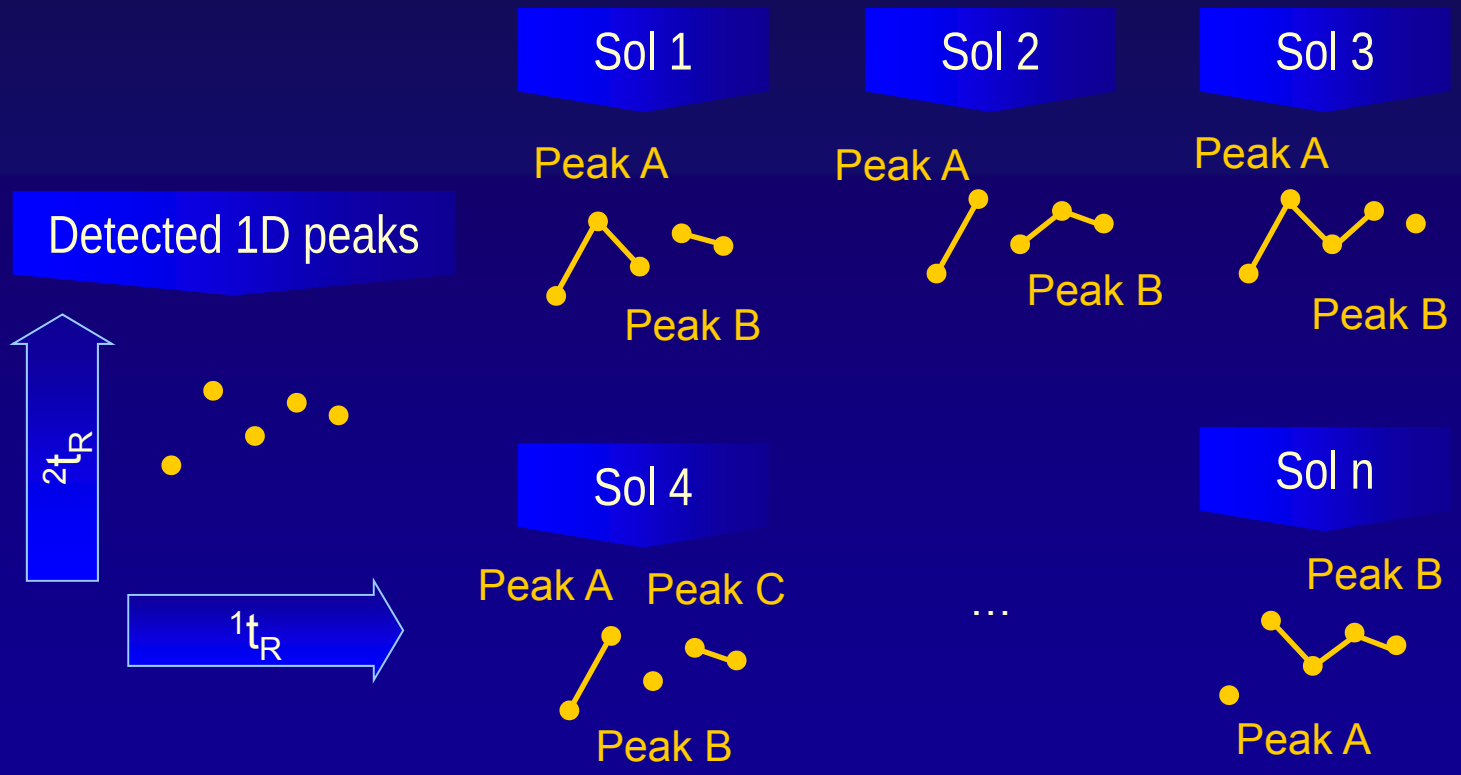
In general, any group of 1D peaks may exhibit x possibilities of arrangement in 2D peaks that do not violate the rules of unimodality and $^2t_R < T$ (tolerance criterion)!!!

Example I: peak detection in two-dimensional chromatography

Example I

Step 1

Let's consider all possible solutions of peak arrangement

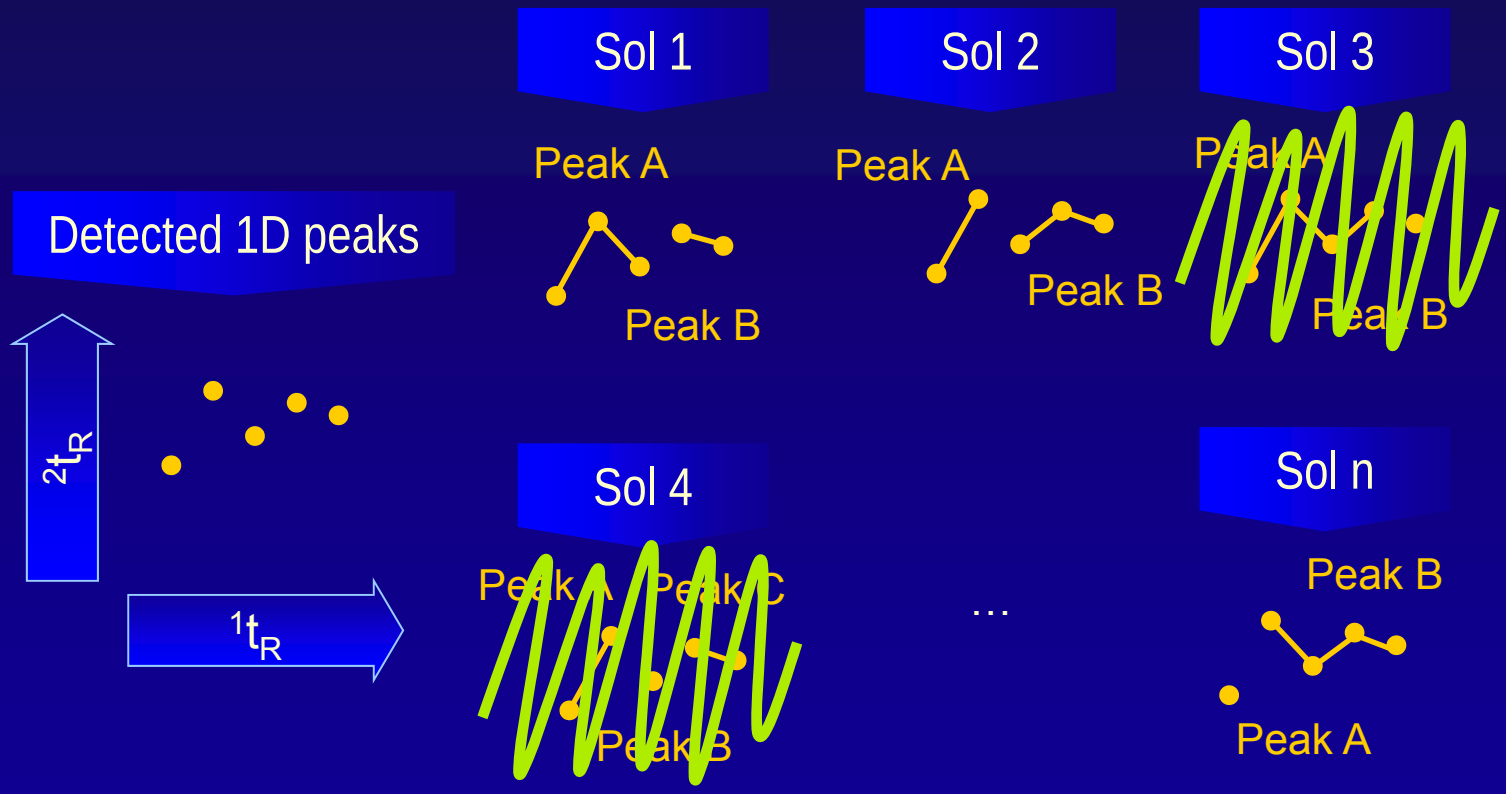


Example I: peak detection in two-dimensional chromatography

Example I

step 2

Discard those solutions that violate the unimodality criterion. Discard also those solutions that imply a too fragmented chromatographic peak.



Example I: peak detection in two-dimensional chromatography

Example I

Step 3

Apply the Bayes theorem to calculate the probability of each solution

Hypothesis

Sol 1

Peak A

Peak B

H_1

Sol 2

Peak A

Peak B

H_2

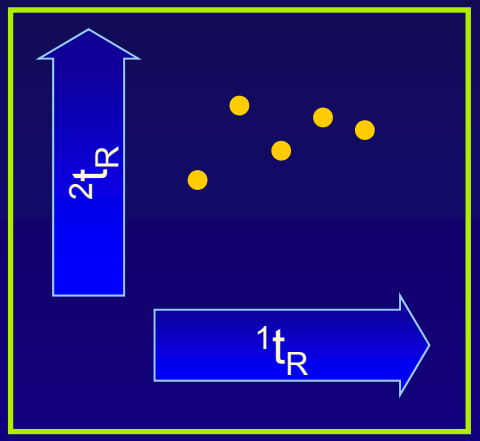
⋮

Peak B

Peak A

Sol n

H_n



$$p(H_i|\mathbf{D}) = p(\mathbf{D}|H_i)\frac{p(H_i)}{p(\boldsymbol{D})}$$

Example I: peak detection in two-dimensional chromatography

Example I

Step 3

Apply the Bayes theorem to calculate the probability of each solution

Hypothesis

Sol 1

Peak A

Peak B

H_1

Sol 2

Peak A

Peak B

H_2

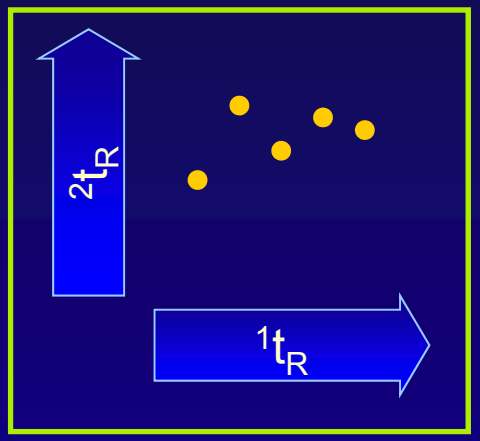
⋮

Peak B

Peak A

Sol n

H_n



$$p(H_i|\mathbf{D}) = p(\mathbf{D}|H_i)\frac{p(H_i)}{p(\boldsymbol{D})}$$

Example I: peak detection in two-dimensional chromatography

Example I

$$p(H_i|\mathbf{D}) = p(\mathbf{D}|H_i)\frac{p(H_i)}{p(\boldsymbol{D})}$$

$$p(H_i|\mathbf{D}) \propto p(\mathbf{D}|H_i)p(H_i)$$

$$p(H_i|\mathbf{D}) \propto p(\mathbf{D}|H_i)$$

I'm interested only in a relative value of $p(H_n|\boldsymbol{D})$

All the priors have the same probability

Example I: peak detection in two-dimensional chromatography

Example I

$$p(\boldsymbol{D}|Hi) = \int \dots \int p(\boldsymbol{D}|\phi_j, {}^1s_j, A_j, Hi)\, p(\phi)p({}^1s)p(A)\, d\phi\, d{}^1s\, dA$$

- ϕ_j: Peak phase
- ${}^1s_j$: 1st dimension peak width
- A_j: Total peak area
- $p(\phi)p({}^1s)p(A)$: Prior probabilities

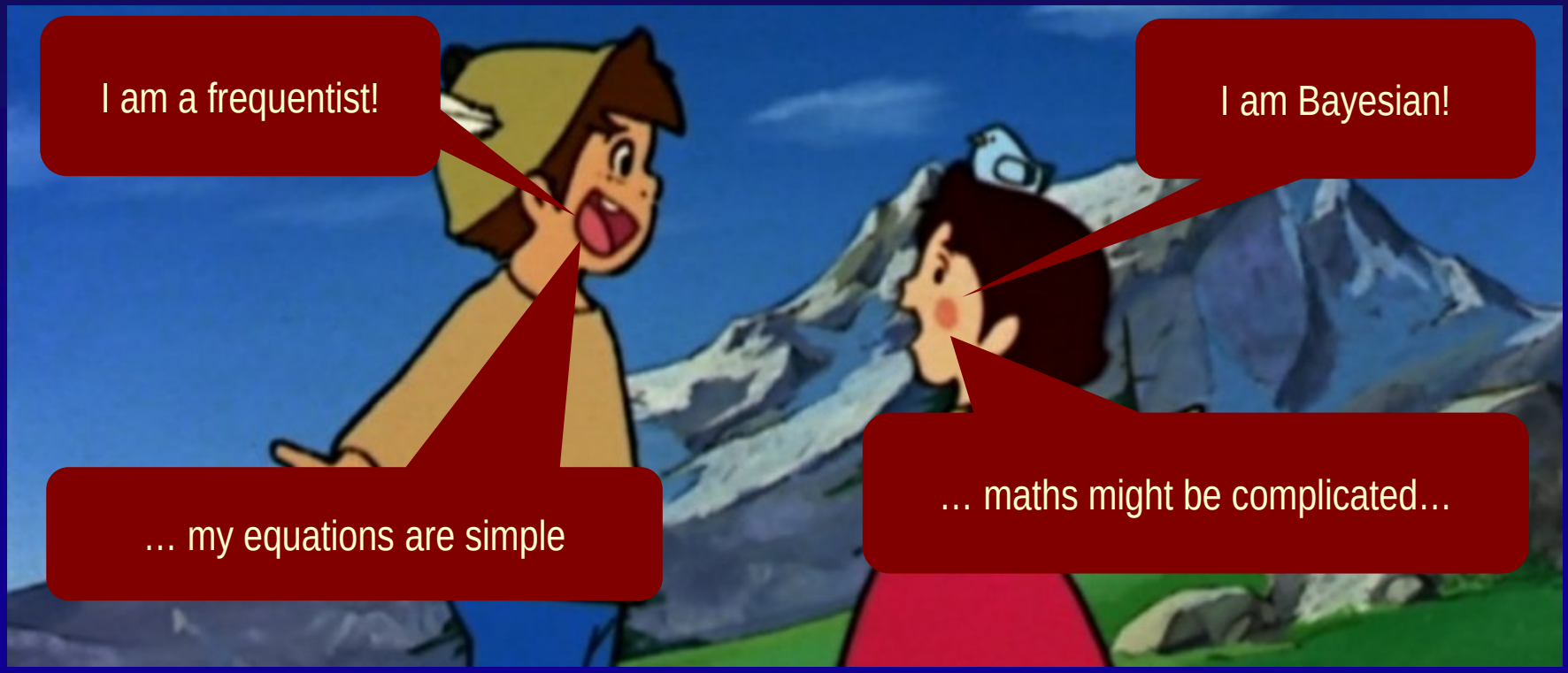
How much does your 1D peak profile look like a peak?

Are the 2nd dimension retention times too far away?

$$p(\boldsymbol{D}|Hi) = \int \dots \int \prod_{j=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{y_j - Aexp\left[-\frac{1}{2}\left(\frac{(\phi + j)m}{{}^1\sigma}\right)^2\right]}{\sigma_y}\right)^2\right] \frac{1}{\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{{}^2tr_j - \overline{{}^2tr}}{\sigma_{2tr}}\right)^2\right] p(\phi)p({}^1s)p(A)\, d\phi\, d{}^1s\, dA$$

Frequentist vs. Bayesian

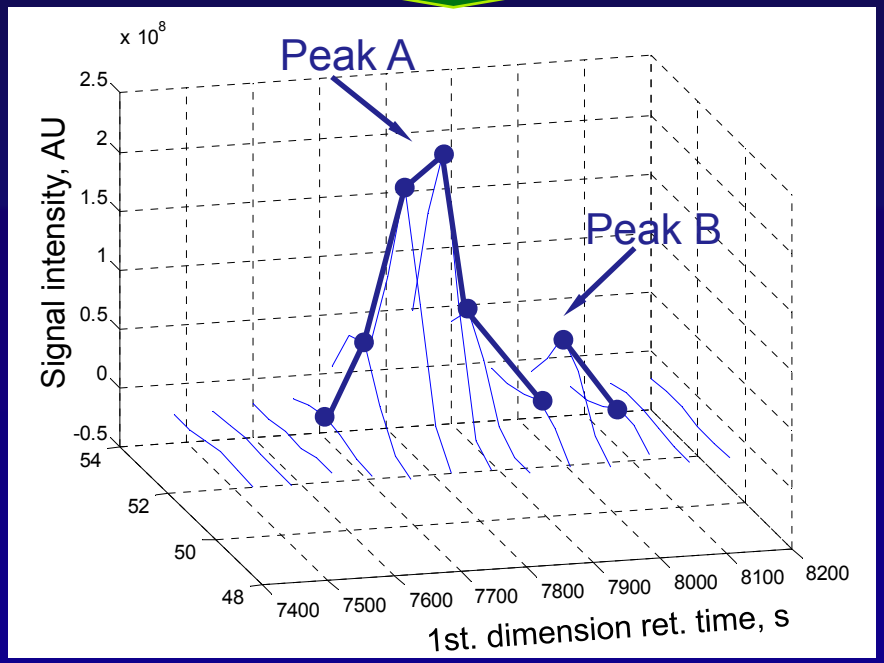
Lesson 3



Example I: peak detection in two-dimensional chromatography

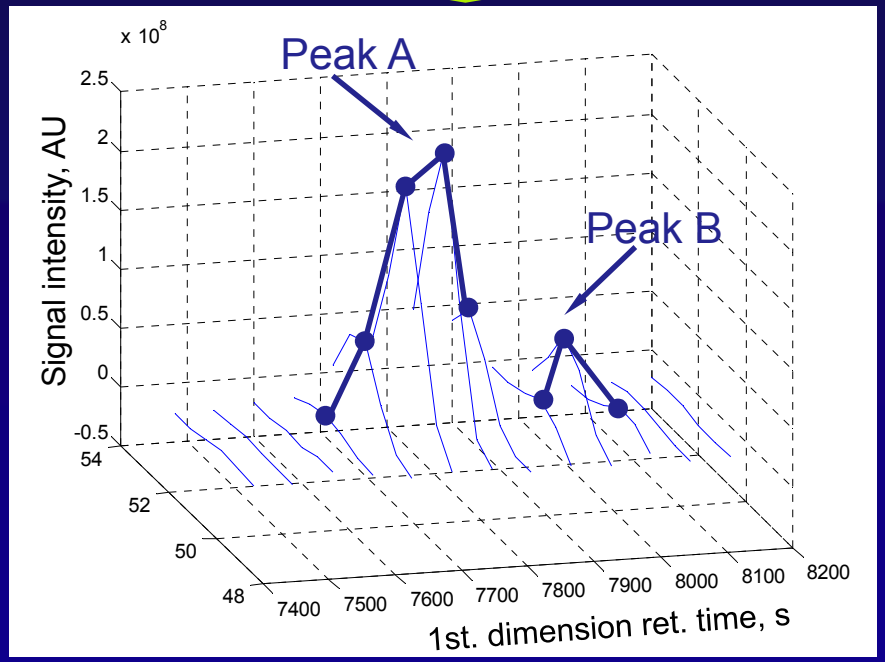
Example I

Possibility 1



Posterior probabilty = 0.51

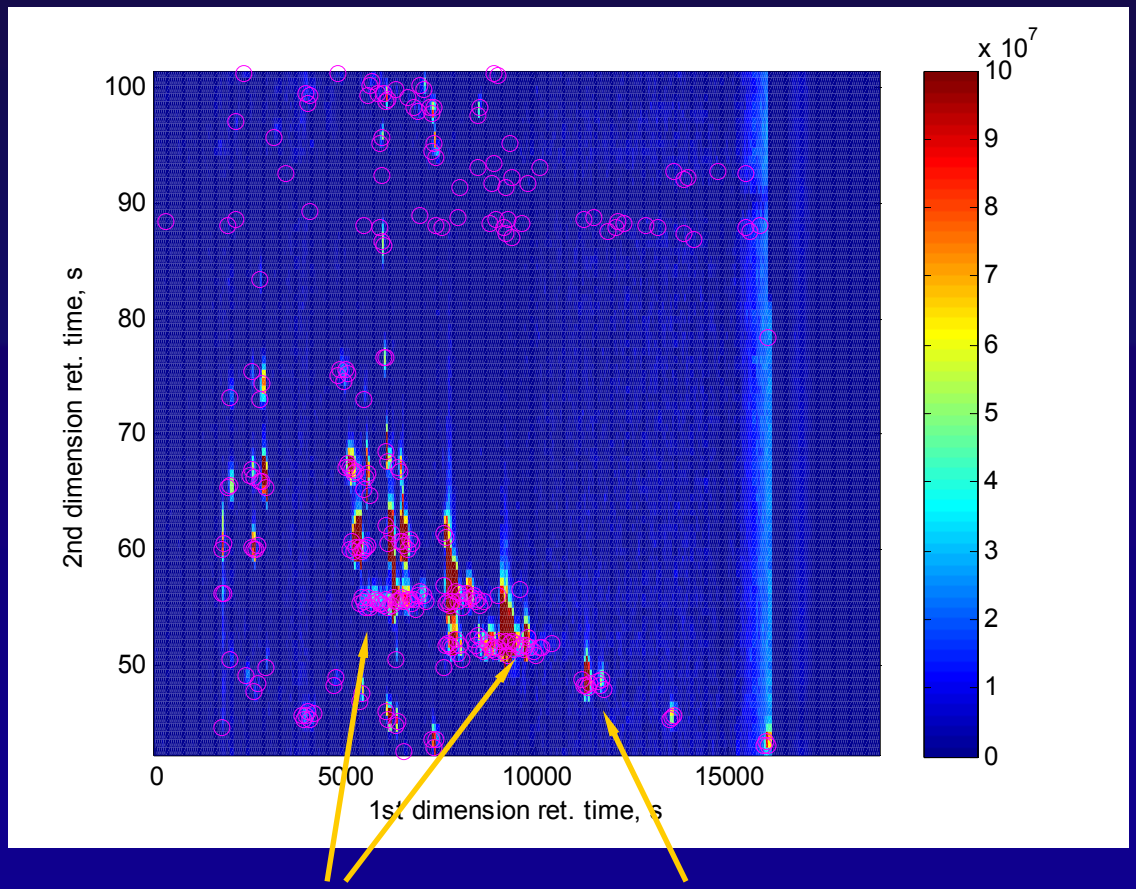
Possibility 2



Posterior probability = 0.49

Example I: peak detection in two-dimensional chromatography

Example I



A map of likelihoods can be constructed, leading to the most probable 2D peak arrangement given the information that the chromatographer has at the moment

Frequentist vs. Bayesian

Lesson 4

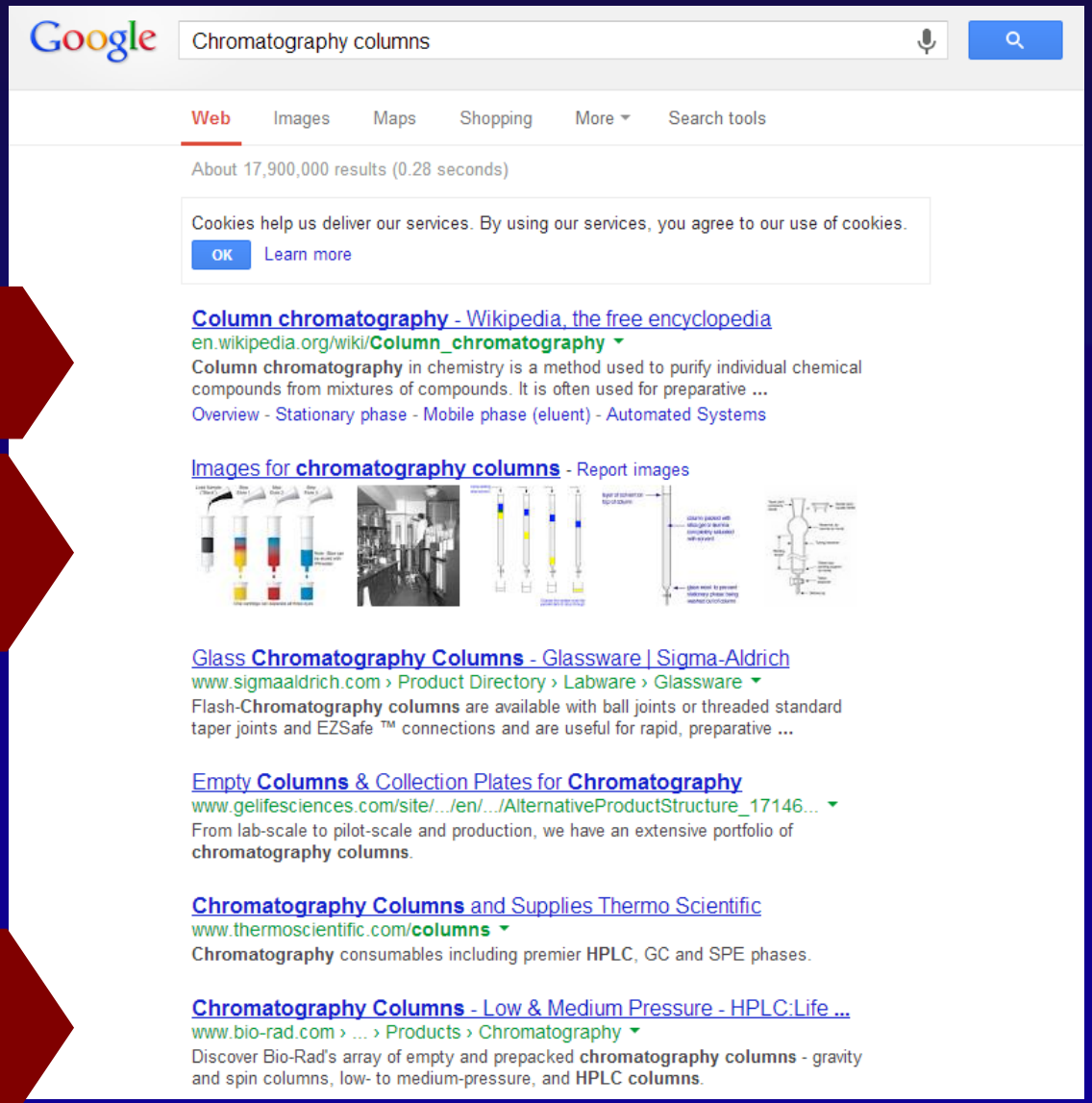
Example I: peak detection in two-dimensional chromatography

Example I

Arrangement H_1

Arrangement H_2

…



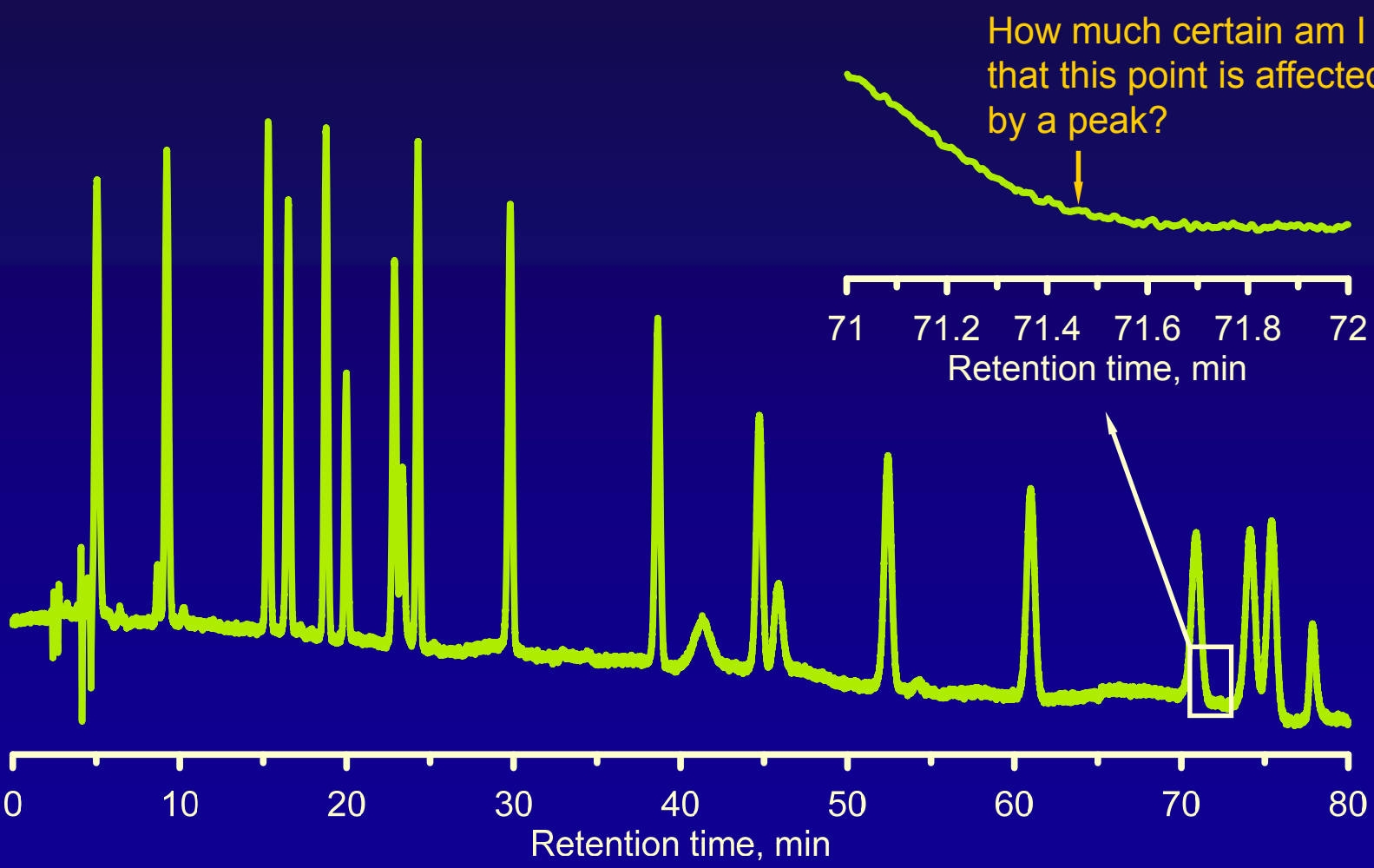
Example I: peak detection in two-dimensional chromatography

Example I

Current algorithms

Example II: peak detection in one-dimensional chromatography

Example II



M. Lopatka, G. Vivó-Truyols, M. Sjerps, Anal. Chim. Acta, 817 (2014) 9-16

Example II: peak detection in one-dimensional chromatography

Example II

Posterior odds = Likelihood ratio x Prior odds

$$\frac{p(H_0|\mathrm{D})}{p(H_1|\mathrm{D})} = \frac{p(\mathrm{D}|H_0)}{p(\mathrm{D}|H_1)} \times \frac{p(H_0)}{p(H_1)}$$

The point IS NOT affected by a peak

The point IS affected by a peak

The data is only used to "update" our prior probability on a situation (but the decision is not "taken" by the algorithm)

M. Lopatka, G. Vivó-Truyols, M. Sjerps, Anal. Chim. Acta, 817 (2014) 9-16

Example II: peak detection in one-dimensional chromatography

Example II

Posterior odds

$$\frac{p(H_0|\mathrm{D})}{p(H_1|\mathrm{D})}$$

=

Likelihood ratio

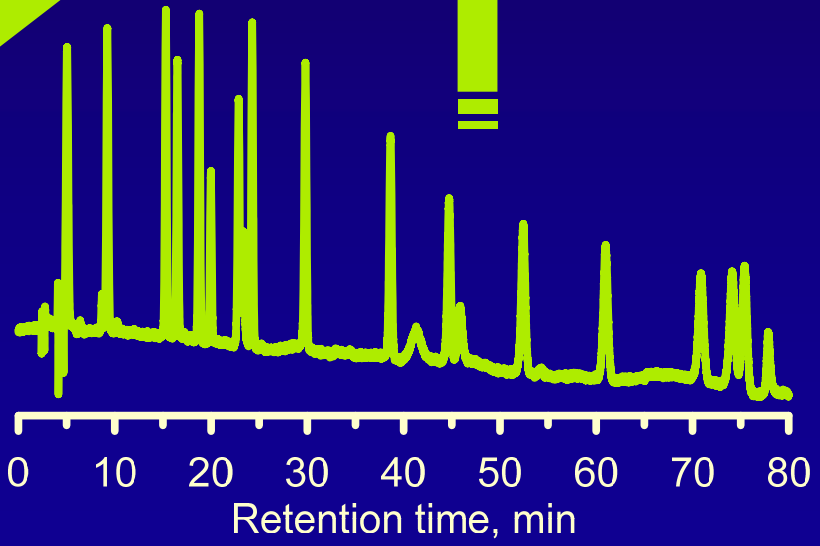
$$\frac{p(\mathrm{D}|H_0)}{p(\mathrm{D}|H_1)}$$

x

Prior odds

$$\frac{p(H_0)}{p(H_1)}$$

Information about possible band broadenings, peak heights, peak asymmetry, etc.



J.M. Davis and J.C. Glddings, Statistical overlap theory, Anal. Chem., 55 (1983).

M. Lopatka, G. Vivó-Truyols, M. Sjerps, Anal. Chim. Acta, 817 (2014) 9-16

Example II: peak detection in one-dimensional chromatography

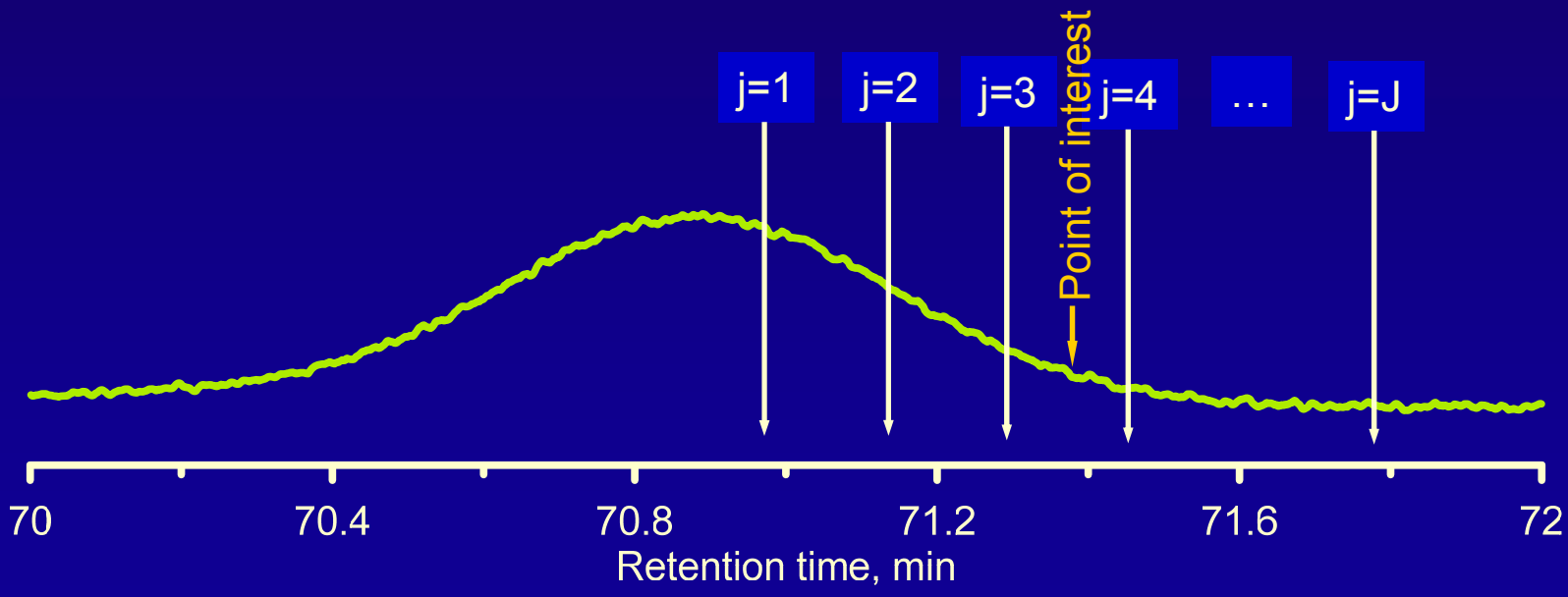
Example II

$$p(\mathrm{D}|H_1) = \sum_{i=1}^{np} \sum_{j=1}^{C_i^J} p(D|i,j,H_1)\, p(j_1,\ldots,j_i|H_1,i)\, p(i|H_1)$$

Assume Gaussian noise

Position(s) of each peak

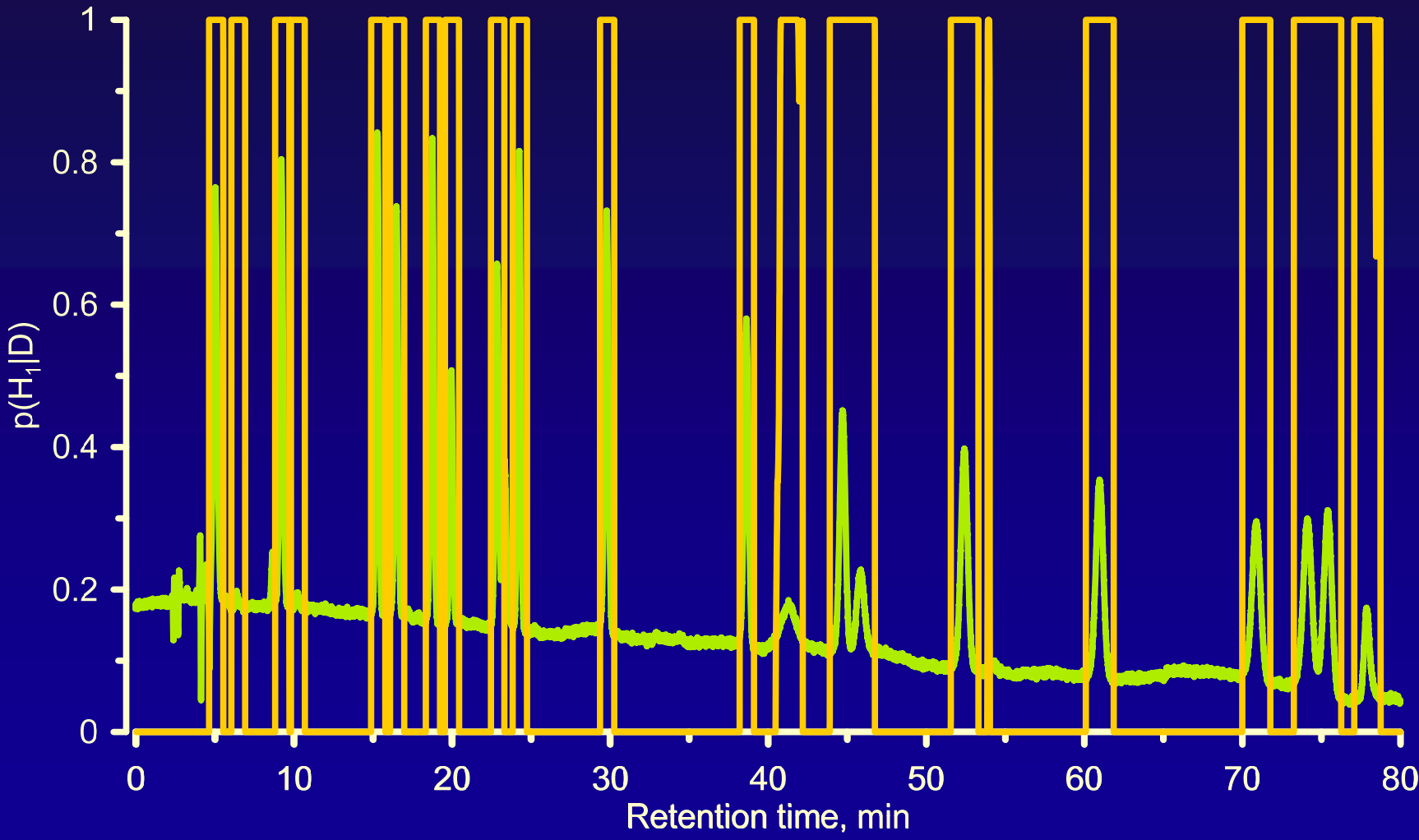
Number of peaks present in the vicinity of the point



M. Lopatka, G. Vivó-Truyols, M. Sjerps, Anal. Chim. Acta, 817 (2014) 9-16

Example II: peak detection in one-dimensional chromatography

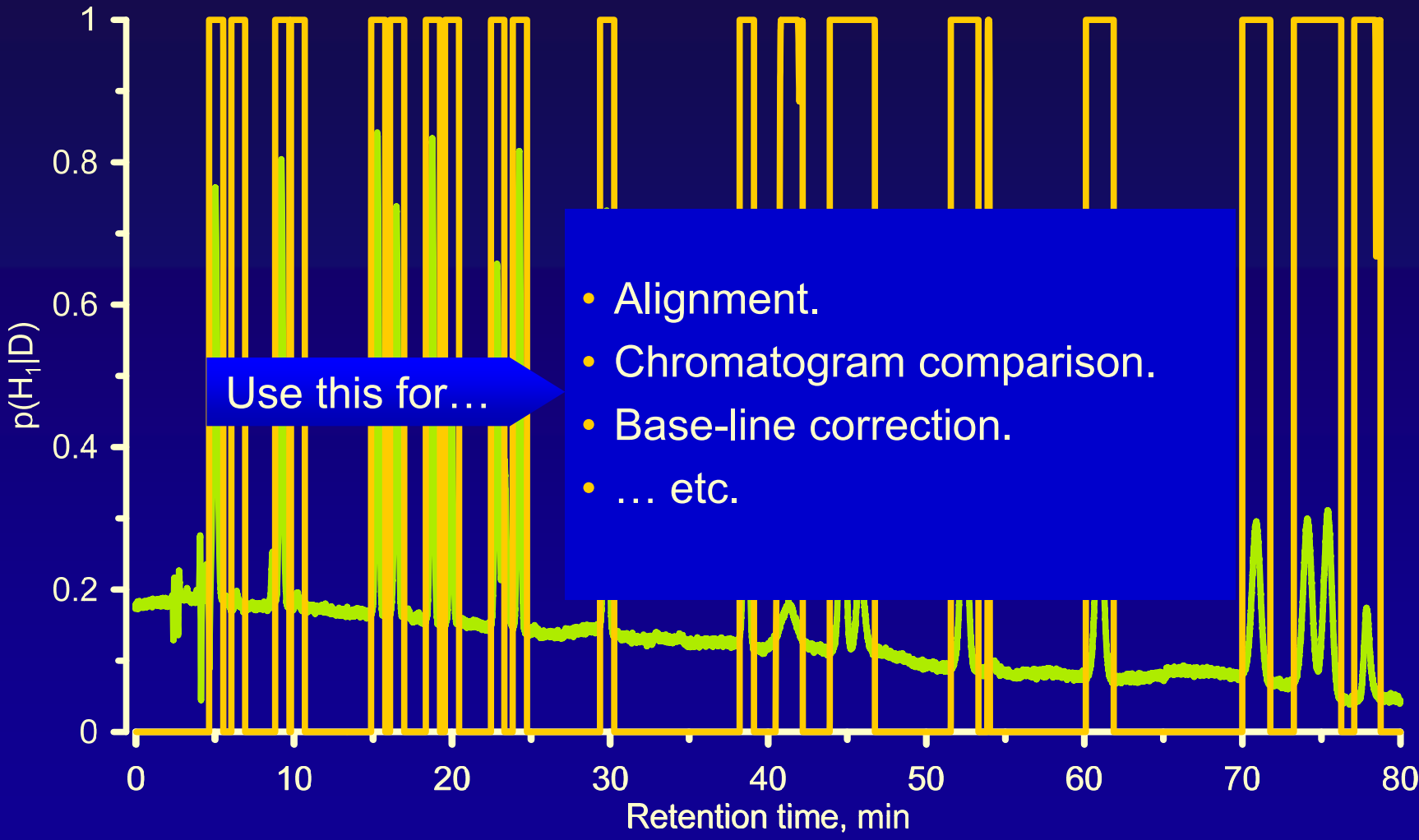
Example II



M. Lopatka, G. Vivó-Truyols, M. Sjerps, Anal. Chim. Acta, 817 (2014) 9-16

Example II: peak detection in one-dimensional chromatography

Example II



Use this for…

- Alignment.
- Chromatogram comparison.
- Base-line correction.
- … etc.

M. Lopatka, G. Vivó-Truyols, M. Sjerps, Anal. Chim. Acta, 817 (2014) 9-16

Frequentist vs. Bayesian

Lesson 5

Example III: screening in forensic toxicology with LC-MS

The problem…

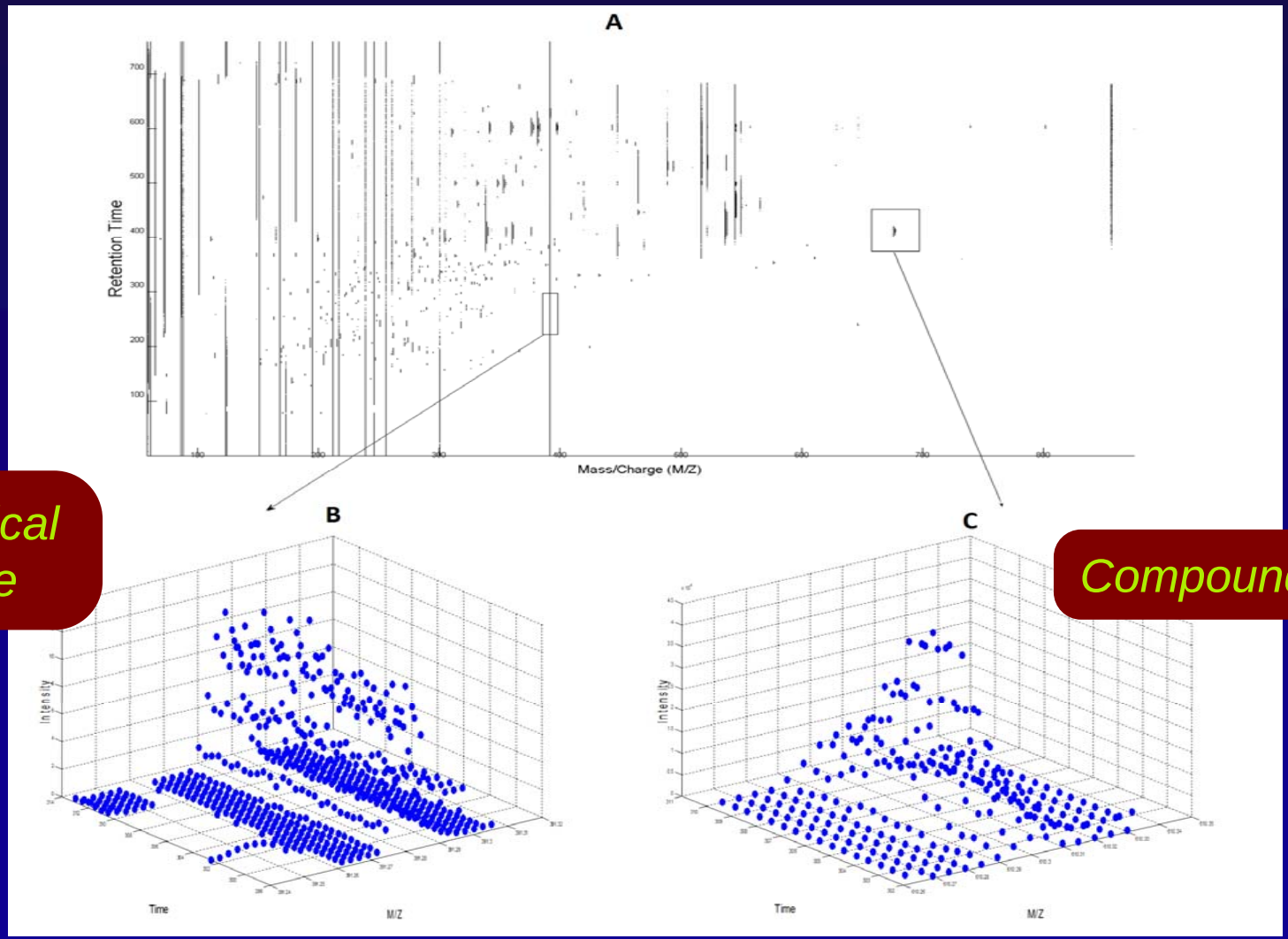
… in a forensic lab, we are intrested in pre-screening the presence/abssence of a fixed list of (~500) compounds using LC-MS

We want a probabilistic value about the presence/absence of a compound, not a final result!

M. Woldegebriel, G. Vivó-Truyols, J. Chromatogr. A, in preparation

Example III: screening in forensic toxicology with LC-MS

Example III



Chemical noise

Compound

M. Woldegebriel, G. Vivó-Truyols, J. Chromatogr. A, in preparation

Example III: screening in forensic toxicology with LC-MS

Example III

Posterior odds = Likelihood ratio x Prior odds

$$\frac{p(H_0|\mathrm{D})}{p(H_1|\mathrm{D})} = \frac{p(\mathrm{D}|H_0)}{p(\mathrm{D}|H_1)} \times \frac{p(H_0)}{p(H_1)}$$

The peak is not present

The peak is present

$p(\mathrm{D}|H_1)$ → Marginalize

This contains LC and MS information (and includes isotopic information)

- The presence/absence of chemical noise
- The possible retention times of the peak (with shifts)
- The possible positions of the MS signal (detector uncertainty)

M. Woldegebriel, G. Vivó-Truyols, J. Chromtogr. A, in preparation

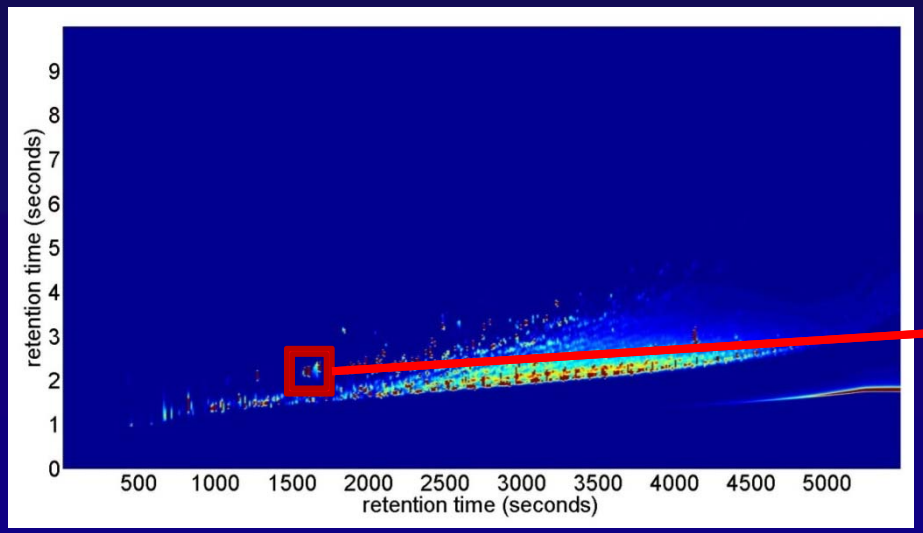
Frequentist vs. Bayesian

Lesson 6

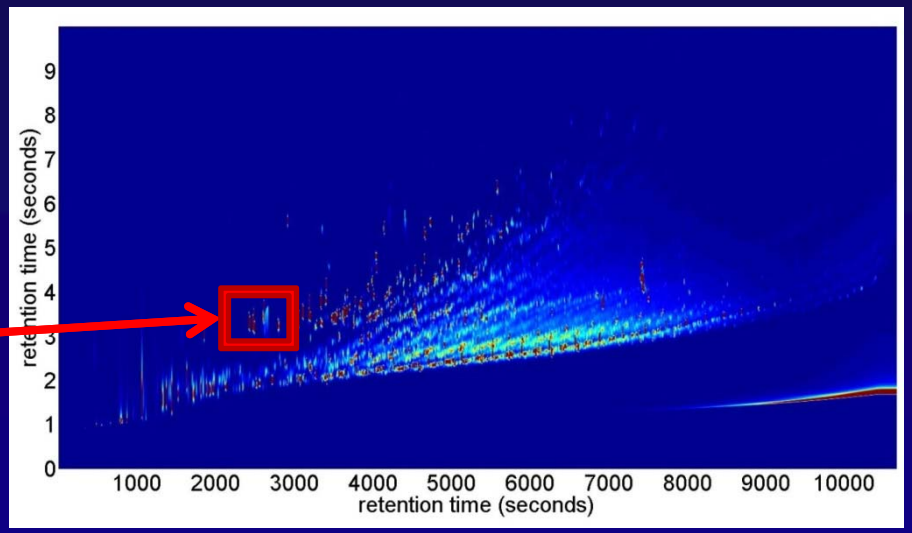
Example IV: Peak tracking in GCxGC

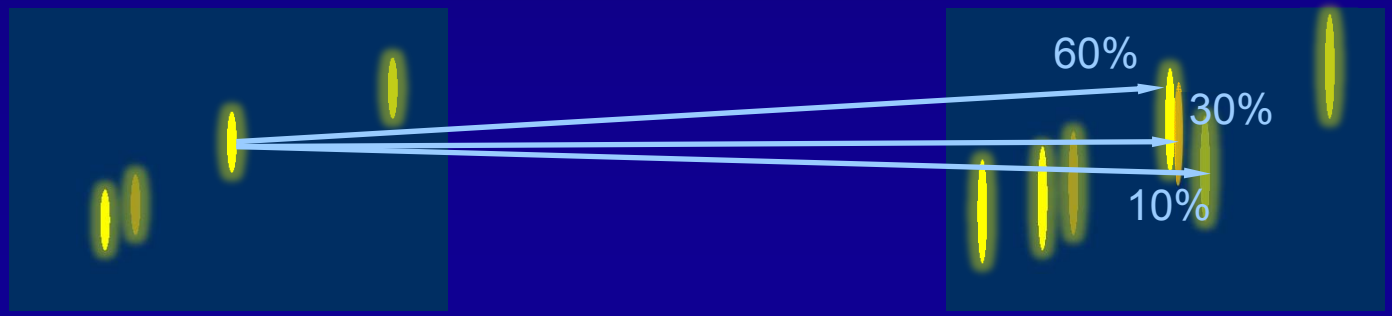
Example IV

GCxGC in condition 1



GCxGC in condition 2





A. Barcaru, G. Vivó-Truyols, J. Chromatogr. A, in preparation

Conclusions

- Automation: the data-analysis part doesn't involve a decision → It is just informing the scientist about the probabilities of the different hypothesis being true…
- 2D peak example: Bayesian analysis computes not only the best peak arrangement, but how probable the rest of alternatives are.
- 1D analysis: Robust and objective method, from prior information (e.g. Statistical Overlap Theory) to final information (probability distributions).
- The methodology can be naturally extended for toxicology screening (including MS). It can handle elegantly complex situation (e.g. different number of isotopes, aducts, etc.). We are beating "Mass Hunter" (Agilent).

Acknowledgements

Special thanks to...

- Analytical-chemistry group (University of Amsterdam)
- T. van Beek (Wageningen university, NL), (2D peak detection example)
- COMFOR project (NWO) for funding.
- COAST consortium (NWO + DSM + NFI + Rikilt) for funding.

Thanks for your attention!