IsoDittwald
Law of Localised Mass Spec Fine Structure

Mateusz Łącki* and Anna Gambin

Faculty of Mathematics, Informatics and Mechanics
University of Warsaw, Banacha 2, 02-097 Warszawa, Poland
mateusz.lacki@biol.uw.edu.pl
aniag@mimuw.edu.pl
http://bioputer.mimuw.edu.pl

Abstract. Approximative distributions theory is used to obtain more tractable formulas describing the localised fine structure of isotopic peaks. We present a new method for calculating localised fine structure isotopic peaks based on the above-mentioned approximations. *abstract* environment.

Keywords: Isotopic Fine Structure, Poisson Approximation, Little Sexy Fox

1 Introduction

There are many reasons why mass spectrometry analysis is hard. It is hard in that there are potentially many many sources of interferences that can distort the information about the actual composition of a sample. The study of the nature of these interferences is needed to achieve the goal of making out of mass spectrometers yet more reliable an identification tool.

Part of the noise in the mass to charge domain is innately related to the elements themeselves and stems from the existence of isotopes. It is because of them that a given analyte is represented as a series of peaks, a spectrum, rather that only one peak. The theoretical underpinnings of how to mathematically model the impact of isotopes are already well established, see [18]. The main idea behind the model is to abstract from the exact positionings of the extra neutrons on a particular chemical compound and thus concentrate only on their relative amounts among all atoms of a given element. Assuming that the isotopic configurations are independent and follow the element dependent distribution, one arrives to the conclusion that the correct law describing occurence of different isotopes in a chemical compound is the product of multinomial distributions.

There is one huge problem with that law: together with the growth of molecule one observes an exponential growth in the number of possibile isotope configurations, which precludes their direct enumeration. To solve this problem, different

* Special thanks to Santa Claus.

semplifications were proposed, amounting to different ways of binning configurations together explicitly [4], by hiding them under the guise of Fourier Transform [13], or by ...

However, it is considered of paramount importance in Mass Spectrometry to develop machines with still higher resolution powers and it is very likely that this trend will continue. Even today there are machines that already can distinguish peaks attributed to different configurations with the same number of extra neutrons.

Add Olson and others but Olson above all.

Here we propose to approach the problem of fine structure so that it overcomes the shortcomings of the aggregate model, as used in [4]. That particular model bins together configurations having the same number of extra neutrons distributed on different atoms. For instance, if one considers water molecule H_2O, the model would glue together configuration with one extra neutron only on the first hydrogen togeter with that having it on the second together with that on oxygen atom. We devise an algorithm to deaggregate these probability clusters. We call the peaks obtained via that algorithm a *localised fine structure*.

What motivates the solution to this problem is a search for better molecule fingerprints. The development of new mass spectrometers capable of distinguishing differences in masses of neutrons is proceeding at a vigorous pace. Soon, scientists will face the need of more detailed models than those abstracting from mass defects. It is also common for chemists to search for presence of specific substance in the sample. Usually this is done by looking at some highly specific range in the mass domain of the gathered spectra. Our model provides deeper insight to what might happen while focussing on that particular bit of collected data: conditioning on configurations with the same number of extra neutrons translates directly into focussing in a specific region of the mass-to-charge domain.

The algorithm assumes that one can easily find a peak not far from the most probable one and that the distribution is close to what one would call unimodal[1] and that the most of distributions probability lies in a rather small neighbourhood of the mode. Both the guess about the starting point and the way the neighbourhood gets explored depend on the Poisson approximation to the distribution under study. To our best knowledge this type of approximation have not yet been used for algorithmic purposes. It has been used however in the context of proteomic and peptide research: in [3] it is being used for high throughput protein identification; its use was revalidated in [17] in case of peptides.

We also observe that the use of Poisson approximation gives a theoretical explanation for the equatransneutronic binning used in [12] and actually helps deaggregating results obtained using that approach as well.

Diophantine equations.

[1] We provide a precise definition of unimodality for discrete probability distributions in Section ...

2 Approximations

By an isotopic configuration we understand information about the number of different isotopes in the sample. For the purpose of simplicity, we focus here on chemical compounds composed of carbon, hydrogen, nitrogen, oxygen, and sulfur; still, the results of this section generalize to any compounds whatsoever. Thus, we concentrate on compounds like $\mathrm{C_cH_hO_oN_nS_s}$, where the low case letters describe the numbers of atoms of particular element type. Among such compounds one can already find peptides and proteins. An isotopic configuration could be represented by an extended empirical formula,

$$ {}^{12}\mathrm{C}_{\mathrm{c}_0}\ {}^{13}\mathrm{C}_{\mathrm{c}_1}\ {}^{1}\mathrm{H}_{\mathrm{h}_0}\ {}^{2}\mathrm{H}_{\mathrm{h}_1}\ {}^{14}\mathrm{N}_{\mathrm{n}_0}\ {}^{15}\mathrm{N}_{\mathrm{n}_1}\ {}^{16}\mathrm{O}_{\mathrm{o}_0}\ {}^{17}\mathrm{O}_{\mathrm{o}_1}\ {}^{18}\mathrm{O}_{\mathrm{o}_2}\ {}^{32}\mathrm{S}_{\mathrm{s}_0}\ {}^{33}\mathrm{S}_{\mathrm{s}_1}\ {}^{34}\mathrm{S}_{\mathrm{s}_2}\ {}^{36}\mathrm{S}_{\mathrm{s}_4}. \quad (1) $$

In the above representation, small letters with indices represent counts of different atoms with indices displaying the number of additional neutrons an isotope has with respect to the lighest possible isotopic variant.

Rather than (1), we shall be using an equivalent probabilistic notation, treating upper case letters, like ${}^{12}\mathrm{C}$, as random variables and considering small case letters, c_0, to be their realisations. An expression like $A = \{{}^{13}\mathrm{C} = \mathrm{c}_1,\ {}^{2}\mathrm{H} = \mathrm{h}_1\}$ is shorthand for saying: let us focus on all configurations (1) that have c_1 heavy carbons and h_1 deuters in total.

Following [10], one assumes that the law of vector

$$ \left({}^{12}\mathrm{C},\ {}^{13}\mathrm{C},\ {}^{1}\mathrm{H},\ {}^{2}\mathrm{H},\ {}^{14}\mathrm{N},\ {}^{15}\mathrm{N},\ {}^{16}\mathrm{O},\ {}^{17}\mathrm{O},\ {}^{18}\mathrm{O},\ {}^{32}\mathrm{S},\ {}^{33}\mathrm{S},\ {}^{34}\mathrm{S},\ {}^{36}\mathrm{S}\right), \quad (2) $$

given $\mathrm{C_cH_hO_oN_nS_s}$, is a product of independent multinomial distributions,

$$ \mathbb{M} = \mathrm{Multi}\Big(\mathbb{P}({}^{12}\mathrm{C}), \mathbb{P}({}^{13}\mathrm{C}); c\Big) \otimes \cdots \otimes \mathrm{Multi}\Big(\mathbb{P}({}^{32}\mathrm{S}), \mathbb{P}({}^{33}\mathrm{S}), \mathbb{P}({}^{34}\mathrm{S}), \mathbb{P}({}^{36}\mathrm{S}); s\Big), \quad (3) $$

where the probabilities of observing particular isotopes, $\mathbb{P}({}^{12}\mathrm{C}), \ldots, \mathbb{P}({}^{36}\mathrm{S})$, are established in independent experiments[2]. For instance, the probability of a given carbons configuration $(\mathrm{c}_0, \mathrm{c}_1)$ equals

$$ \mathrm{Multi}\left(\mathbb{P}({}^{12}\mathrm{C}), \mathbb{P}({}^{13}\mathrm{C}); c\right)\Big((\mathrm{c}_0, \mathrm{c}_1)\Big) = \binom{\mathrm{c}}{\mathrm{c}_0, \mathrm{c}_1} \mathbb{P}({}^{12}\mathrm{C})^{\mathrm{c}_0} \mathbb{P}({}^{13}\mathrm{C})^{\mathrm{c}_1} $$

and it should be multiplied by similar expression for hydrogen, nitrogen, oxygen and sulfur to obtain probability for expression like (1).

Observe, that given $\mathrm{C_cH_hO_oN_nS_s}$, part of the information in (2) is redundant and can be shortened by neglecting counts of the lightest isotope variants, leaving us with

$$ \left({}^{13}\mathrm{C},\ {}^{2}\mathrm{H},\ {}^{15}\mathrm{N},\ {}^{17}\mathrm{O},\ {}^{18}\mathrm{O},\ {}^{33}\mathrm{S},\ {}^{34}\mathrm{S},\ {}^{36}\mathrm{S}\right). \quad (4) $$

Missing therms can be retrieved from relationships ${}^{12}\mathrm{C} + {}^{13}\mathrm{C} = \mathrm{c}$, ${}^{1}\mathrm{H} + {}^{2}\mathrm{H} = \mathrm{h}$, and so on, that occur with probability one.

[2] Consult Table 1 for details.

Definition 1 *We call the set of configurations*

$$LFS_K = \left\{ {}^{13}C + {}^{2}H + {}^{15}N + {}^{17}O + 2\times{}^{18}O + {}^{33}S + 2\times{}^{34}S + 4\times{}^{36}S = K \right\} \tag{5}$$

a localised fine structure with K extra neutrons.

The reason for numbers 2 and 4 appearing above is that ^{18}O and ^{34}S have two additional neutrons, and ^{36}S – four; confront Table 1.

The problem of finding the cardinal number of LFS_K is also known as the money exchange problem. In general, enumeration of all elements of LFS_K corresponds to finding all integer solutions $(x_1, \ldots, x_k)$ to a *Linear Diophantine Equation*

$$a_1 x_1 + \cdots + a_k x_k = K, \tag{6}$$

where $(a_1, \ldots, a_k)$ are integer coefficients. It is known, see [1], that the number of solutions to (6) is approximately $\frac{K^{k-1}}{(k-1)!a_1 \ldots a_k}$ if only the greatest common divisor of the equation's parameters is equal to one. It is hard to imagine an organic substance without carbon, show that this proxy in our applications always holds. This is a big problem, but there is a natural ordering on solutions *via* probability (3) and we concentrate on finding only the most probable configurations.

Problem 1 *For a given K, find a small set $B \subset LFS_K$ of configurations s.t.*

$$\mathbb{M}_K(B) := \frac{\mathbb{M}(B)}{\mathbb{M}(LFS_K)} \approx 1\,. \tag{7}$$

In statistical terms, we are interested in approximating the critical set of the distribution on configurations that make part of a given *localised fine structure*.

Why should one study law described by (7) in the first place? The mass of all configurations in a *localised fine structure* can be approximated at first by the monoisotopic mass shifted by the the number of extra neutrons times one Dalton. By doing that, we assume that all extra neutrons weight exactly one Dalton. That assumption is used in many algorithms, as BRAIN [4], ...

Add more softwares

That is a good approximation, since the difference between masses are much smaller than one Dalton for different elements, see [8]. If that was the case, one could not discern among configurations gathered in LFS_K. One might be thought interested in studying yet more precise molecule's fingerprint. This can be achieved by looking precisely at LFS_K. All of configurations in LFS_K are localised in the same region of the mass to charge domain, being the proximity of the above-mentioned approximation. This has additional advantages resulting from the construction of machines, as described in the last section of the article.

Wait for Dirk or Frederik's answer on the resolution being different in different mass to charge regions.

To solve Problem 1 we proceed by approximating

Change exposition.

the *full distribution* (3) by a product of Poisson laws.

Lemma 1. *Consider a multionomial distribution*

$$\mathrm{Multi}^{[n]}(r_0, r_1, \ldots, r_w; n) = \binom{n}{r_0, r_1, \ldots, r_w} p_{0,n}^{r_0} p_{1,n}^{r_1} \cdots p_{w,n}^{r_w}.$$

If $\lim_{n\to\infty} np_{k,n} = \lambda_k$ *exist for* $k = 1, \ldots, w$ *then*

$$\mathrm{Multi}^{[n]} \rightharpoonup \delta_\infty \otimes \mathrm{Poiss}(\lambda_1) \otimes \cdots \otimes \mathrm{Poiss}(\lambda_w), \tag{8}$$

where δ_∞ *is a measure concentrated on* ∞ *and* Poiss *is the Poisson distribution,*

$$\mathrm{Poiss}(\lambda)(k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

The proof is well known in the literature and we omit it[3].

Note that the approximation assumes that we enlarge the number of trials in the multinomial distribution to infinity. In the context of our model this would correspond to infinite enlargement of the compound so that only the lightest isotopes of differenet elements would appear infinitely often, other taking any finite value. Thus, the Poisson approximation enlarges the state space of the problem to configurations that are nonphysical. The interpretational shortcomings are overweighted however by the emerging independence of the numbers of isotope counts.

We tackle the problem of infinite numbres of lightest isotopes in the following way: we assume, that the configurations to which we can prescribe the approximative distributions are simply the counts of the heavier isotopes and call that a reduced configuration. In the example studied in this paper it amounts to

$$^{13}\mathrm{C}_{\mathrm{c}_1}\ ^{2}\mathrm{H}_{\mathrm{h}_1}\ ^{15}\mathrm{N}_{\mathrm{n}_1}\ ^{17}\mathrm{O}_{\mathrm{o}_1}\ ^{18}\mathrm{O}_{\mathrm{o}_2}\ ^{33}\mathrm{S}_{\mathrm{s}_1}\ ^{34}\mathrm{S}_{\mathrm{s}_2}\ ^{36}\mathrm{S}_{\mathrm{s}_4}. \tag{9}$$

The reduction is a common approach to the problem; confront [7]. Observe, that for a reduced isotopic configuration the constraint (**??**) is still a valid one, being expressed only in terms of numbers of not the lightest isotopes.

Another question worth addressing what should be chosen for n, while applying Lemma 1. It is an important question, since for cerain values n the approximation works better. A detailed description of this phenomenon can be found in [14]. Since we approximate each multinomial distribution in (3) it is natural to consider more than one value: one should look at the numbers of different elements in the chemical compound, i.e. on the empirical formula $\mathrm{C_cH_hO_oN_nS_s}$. The bigger the minimal number atoms, the better the approximation should be[4]. In case of peptides and proteins, Senko et al. [16] introduced the concept of avergine, an averaged chain of m amino acids, with empirical formula

$$\mathrm{C}_{\lfloor m\times 4.9384\rfloor}\mathrm{H}_{\lfloor m\times 7.7583\rfloor}\mathrm{O}_{\lfloor m\times 1.4773\rfloor}\mathrm{N}_{\lfloor m\times 1.3577\rfloor}\mathrm{S}_{\lfloor m\times 0.0417\rfloor},$$

We infer n's from that relationship while using Lemma 1 for peptides and proteins.

Finally, in the approximation we *calibrate* λ's setting them to be equal to average numbers of isotopes using original *full distribution* (3). For instance, for

[3] It makes part of common knowledge: mathematicians are more concerned about measuring the quality of this approximation, as in [14].

[4] More precisely: the smaller is the total variance difference between the approximation and the approximated term.

carbon we set $\lambda_{{}^{13}\mathrm{C}} \approx \mathrm{c} \times \mathbb{P}({}^{13}\mathrm{C})$. This is in contrast to the *fitting* approach used in [3, 17], where the means of the approximation are free parameter in an optimisation scheme. These two solutions should not differ too much for larger compounds, for it is known that both the Poisson and Multinomial distributions are concentrated near their modes, see [2].

Hence, the probability that we assign to reduced configuration (9) is equal to

$$\frac{\lambda_{{}^{13}\mathrm{C}}^{\mathrm{c}_1}}{\mathrm{c}_1!}\frac{\lambda_{{}^{2}\mathrm{H}}^{\mathrm{h}_1}}{\mathrm{h}_1!}\frac{\lambda_{{}^{15}\mathrm{N}}^{\mathrm{n}_1}}{\mathrm{n}_1!}\frac{\lambda_{{}^{17}\mathrm{O}}^{\mathrm{o}_1}}{\mathrm{o}_1!}\frac{\lambda_{{}^{33}\mathrm{S}}^{\mathrm{s}_1}}{\mathrm{s}_1!}e^{-\mu}\frac{\lambda_{{}^{18}\mathrm{O}}^{\mathrm{o}_2}}{\mathrm{o}_2!}\frac{\lambda_{{}^{34}\mathrm{S}}^{\mathrm{s}_2}}{\mathrm{s}_2!}e^{-\eta}\frac{\lambda_{{}^{36}\mathrm{S}}^{\mathrm{s}_1}}{\mathrm{s}_1!}e^{-\gamma} \tag{10}$$

where

$$\begin{aligned}\mu &= \lambda_{{}^{13}\mathrm{C}} + \lambda_{{}^{2}\mathrm{H}} + \lambda_{{}^{15}\mathrm{N}} + \lambda_{{}^{17}\mathrm{O}} + \lambda_{{}^{33}\mathrm{S}}\\ \eta &= \lambda_{{}^{18}\mathrm{O}} + \lambda_{{}^{34}\mathrm{S}}\\ \gamma &= \lambda_{{}^{36}\mathrm{S}}.\end{aligned}$$

The usefulness of approximation by a product of independent Poisson lies closed formula expression one obtains while conditioning.

Lemma 2. *Suppose we have a collection of m independent Poisson-distributed random variables, $X_i \sim \mathrm{Poiss}(\mu_i)$. Then $X_1, \ldots, X_m$ given that $X_1 + \cdots + X_m = K$ is multinomially distributed,*

$$\Big(X_1, \ldots, X_m | X_1 + \cdots + X_m = K\Big) \sim \mathrm{Multi}\Big(\frac{\mu_1}{\sigma}, \ldots, \frac{\mu_m}{\sigma}; K\Big),$$

where $\sigma = \sum_{i=1}^{m} \mu_i$.

Proof might be found in [11].

Suppose that we concentrated on molecule composed of elements that have only one additional neutron, e.g. $\mathrm{C_cH_hN_n}$. Then, following Lemma 2,

Result 1 *The approximative distribution of the* localised fine structure *with K extra neutrons of $C_cH_hN_n$ is*

$$\mathrm{Multi}\left(\frac{\lambda_{{}^{13}C}}{\mu}, \frac{\lambda_{{}^{2}H}}{\mu}, \frac{\lambda_{{}^{15}N}}{\mu}; K\right). \tag{11}$$

However, it is not yet clear why should it be true that we can approximate the *localised fine structure law* by the Poisson approximation conditional on the set of configurations with the same total number of extra neutrons. What remains to be shown is why the conditioning does not preclude convergence in distribution. This, however, can be easily proved.

Lemma 3. *Let $\mu^{[n]}, \mu$ be discrete measures. If $\mu^{[n]}$ converges in distribution to μ and an event A has non zero probability under any of that measures, $\underset{n}{\forall}\mu^{[n]}(A)\,,\ \mu(A) > 0$, then measures condtional on A, $\mu_A^{[n]} = \frac{\mu^{[n]}}{\mu^{[n]}(A)}$ converge in distribution to $\mu_A = \frac{\mu}{\mu(A)}$.*

The proof is to be found in the appendix.

In our case, $\mu_A^{[n]}$ is the projection of *full distribution* onto the space of reduced configurations, conditioned on the set $A = \{(\mathrm{c}_1, \mathrm{h}_1, \mathrm{n}_1) : \mathrm{c}_1 + \mathrm{h}_1 + \mathrm{n}_1 = K\}$. That would get approximated by (11). Observe that in probabilistic notation

$$A = \left\{ {}^{13}\mathrm{C} + {}^{2}\mathrm{H} + {}^{15}\mathrm{N} = K \right\}.$$

It is valuable to see, how Lemma 2 generalizes while conditioning on a particular *localised fine structure* when the compound is composed out of elements with multiple isotopes. The problem is that the set of configurations with a fixed number of extra neutrons corresponds to a different Diophantine equation: namely the condition defining A might be like (**??**). The Poisson approximation simplifies the calculations of probability assigned to set A. It stems from the following

Be sure that the concept of Diophantine equation is introduced.

Lemma 4. *Suppose we have a collection of m independent Poisson-distributed random variables, $X_i \sim \mathrm{Poiss}(\mu_i)$. Then $X_1 + \cdots + X_m \sim \mathrm{Poiss}(\mu_1 + \cdots + \mu_m)$.*

The proof can be found in [11].

Note that A can be described by sums of three different Poisson variables instead of eight:

$$A = \Big\{ \underbrace{{}^{13}\mathrm{C} + {}^{2}\mathrm{H} + {}^{15}\mathrm{N} + {}^{17}\mathrm{O} + {}^{33}\mathrm{S}}_{G_1} + 2 \times \underbrace{({}^{18}\mathrm{O} + {}^{34}\mathrm{S})}_{G_2} + 4 \times \underbrace{{}^{36}\mathrm{S}}_{G_4} = K \Big\}. \quad (12)$$

where $G_1 \sim \mathrm{Poiss}(\mu)$, $G_2 \sim \mathrm{Poiss}(\mu)$, and $G_4 \sim \mathrm{Poiss}(\mu)$. There is a strict link between random variables G_i and the concept of equatransneutronic groups described in [12]: it is equal to the total number of atoms in a compound bearing additional i neutrons. Also, let us define three numbers

$$\begin{aligned} x &= \mathrm{c}_1 + \mathrm{h}_1 + \mathrm{n}_1 + \mathrm{o}_1 + \mathrm{s}_1, \\ y &= \mathrm{o}_2 + \mathrm{s}_2, \\ z &= \mathrm{s}_4. \end{aligned}$$

In [12] they are encoded by k_1, k_2, and k_4; also, $d_{G_i} = i$. Then it is true that

Result 2 *The approximative* fine structure law *with K additional neutrons for $C_c H_h O_o N_n S_s$ is equal to*

$$\mathrm{Multi}\left(x; \frac{\lambda_{{}^{13}C}}{\mu}, \frac{\lambda_{{}^{2}H}}{\mu}, \frac{\lambda_{{}^{15}N}}{\mu}, \frac{\lambda_{{}^{17}O}}{\mu}, \frac{\lambda_{{}^{33}S}}{\mu}\right) \otimes \mathrm{Multi}\left(y; \frac{\lambda_{{}^{18}O}}{\eta}, \frac{\lambda_{{}^{34}S}}{\eta}\right) \otimes \mathbb{L}(x, y, z),$$

where

$$\mathbb{L}(x, y, z) = \frac{\frac{\mu^x}{x!}\frac{\eta^y}{y!}\frac{\gamma^z}{z!}}{\sum\limits_{x'+2y'+4z'=K} \frac{\mu^{x'}}{x'!}\frac{\eta^{y'}}{y'!}\frac{\gamma^{z'}}{z'!}}. \quad (13)$$

Thus, Result 2 also provides us with (13) – a natural distribution on the *equatransneutronic configurations*. In general, the elements forming a chemical compound may have different numbers of extra neutrons than $I = \{1, 2, 4\}$. For a general set I formula (13) generalizes to

$$\mathbb{L}(\boldsymbol{k}) = \frac{\prod_{i\in I} \frac{\mu_i^{k_i}}{k_i!}}{\sum\limits_{\{\boldsymbol{k}^*:\sum_{i\in I} ik_i^*=K\}} \prod_{i\in I} \frac{\mu_i^{k_i^*}}{k_i^*!}},$$

where $\boldsymbol{k}$ is an ordered tuple indexed by I. Observe, that in nature at most $\#I \leq 10$, which poses a limit on the complexity of calculating all the values of $\mathbb{L}$. We call $\mathbb{L}$ the *lucky distribution* and note, that is is equivalent to conditioning $\#I$ independent Poisson distributions with different parameters, $\boldsymbol{\mu}$, on the set of solutions to Diophantine equation $\sum_{i\in I} ik_i = K$.

3 Algorithms

4 Discussion and Conclusions

Novelties:

- Greedy approach based on concentration of measure considerations.
- Top-down identification compatibility.

Advantage: Say that it suits extremely well a top-down identification strategy.

One could say that Problem 1 is not the one worth studying. Assuming that the machines are already advanced enough to resolve peaks that come from different configurations with the same number of neutrons, why should one not pose the following:

Problem 2 *Find a small set C among all possible configurations, s.t.* $\mathbb{M}(C) \approx 1$.

Write about the gready approach we use to get the peaks.

Observe, that such a problem could be solved in a gready manner[5], similar to one described in the previous section, with a naturally arrising candidate for the biggest peak: the direct product of modes of each multinomial model described by (3). However, at least in case of Time of Flight analyzers, there is one potential advantage of studying the LFS_K configurations over the ones gathered in the critical set C: it is known, that in these instruments the resolution depends on the mass of analyte, see [6]. It is more difficult to differentiate correctly molecules with similar masses, when both of them are big. There is no guarantee that while proceeding with calculations of elements of C one would not obtain peaks from different regions in the mass to charge domain. Therefore, some sort of

[5] The development of such an algorithm is already proceeding

additional error term in the resolution would have to be included and it is not straightforward to model it. By studying LFS_K we neglect that sort of problem, because of the localisation in the mass to charge domain.

Observe also, that one might have though of approaching Problem 1 by using the above mentioned algorithm, or, in general, any algorithm that results in obtaining *fine structure* masses, and subsequently, by discarding all configurations with masses outside a prespecified interval. This might, however, tremendously prolongue the algorithm's operating time and additional considerations to bypass this problem would have to be made.

Acknowledgments. We would like to thank Piotr Dittwald for thourough introduction to the problem of fine isotopic structure and many productive brawls over proper definitions. We also thank Dirk Valkenborg for pointing out the existence of the concept of equatransneutronic isotopes. Finally, a huge thanks goes to prof. Alan Rockwood, for the time we spent together discussing the mass spec related issues and also for showing me America.

References

1. Agnarsson, G.: On the sylvester denumerants for general restricted partitions. (154), 49–60 (2002)
2. Bobkov, S., Ledoux, M.: On modified logarithmic sobolev inequalities for bernoulli and poisson measures. Journal of Functional Analysis 156(2), 347 – 365 (1998), http://www.sciencedirect.com/science/article/pii/S0022123697931876
3. Breen, E.J., Hopwood, F.G., Williams, K.L., Wilkins, M.R.: Automatic poisson peak harvesting for high throughput protein identification. Electrophoresis 21(11), 2243–2251 (2000), http://dx.doi.org/10.1002/1522-2683(20000601)21:11<2243::AID-ELPS2243>3.0.CO;2-K
4. Claesen, J., Dittwald, P., Burzykowski, T., Valkenborg, D.: An Efficient Method to Calculate the Aggregated Isotopic Distribution and Exact Center-Masses. Journal of The American Society for Mass Spectrometry 23(4), 753–763 (Apr 2012), http://dx.doi.org/10.1007/s13361-011-0326-2
5. Dugundji, J.: Topology. Allyn and Bacon series in advanced mathematics, Allyn and Bacon, Boston (1966), http://opac.inria.fr/record=b1079460
6. Eidhammer, I., Flikka, K., Martens, L., Mikalsen, S.O.: Computational Methods for Mass Spectrometry Proteomics. Wiley-Interscience (2007)
7. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 1. Wiley (January 1968)
8. Hughey, C.A., Hendrickson, C.L., Rodgers, R.P., Marshall, A.G., Qian, K.: Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. Anal. Chem. 73(19), 4676–4681 (Oct 2001)
9. Kallenberg, O.: Foundations of modern probability. Probability and its applications, Springer, New York (2002)
10. Kienitz, H.: Mass Spectrometry and its Applications to Organic Chemistry. Angewandte Chemie 73(17-18) (1961)
11. Kingman, J.F.C.: Poisson processes, Oxford Studies in Probability, vol. 3. The Clarendon Press Oxford University Press, New York (1993), Oxford Science Publications

12. Olson, M., Yergey, A.: Calculation of the isotope cluster for polypeptides by probability grouping. Journal of the American Society for Mass Spectrometry 20(2), 295–302 (2009), http://dx.doi.org/10.1016/j.jasms.2008.10.007
13. Rockwood, A.L.: Relationship of Fourier transforms to isotope distribution calculations. Rapid Commun. Mass Spectrom. 9(1), 103–105 (jan 1995), http://dx.doi.org/10.1002/rcm.1290090122
14. Roos, B.: On the rate of multivariate poisson convergence. Journal of Multivariate Analysis 69(1), 120 – 134 (1999), http://www.sciencedirect.com/science/article/pii/S0047259X98917894
15. Rosman, K., Taylor, P.: Isotopic Compositions of the Elements 1997. J. Phys. Chem. Ref. Data 27(6) (1998)
16. Senko, M.W., Beu, S.C., McLaffertycor, F.W.: Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. J. Am. Soc. Mass Spectrom. 6(4), 229–233 (Apr 1995)
17. Valkenborg, D., Assam, P., Thomas, G., Krols, L., Kas, K., Burzykowski, T.: Using a poisson approximation to predict the isotopic distribution of sulphur-containing peptides in a peptide-centric proteomic approach. Rapid Communications in Mass Spectrometry 21(20), 3387–3391 (2007), http://www.scopus.com/inward/record.url?eid=2-s2.0-35349002401&partnerID=40&md5=d68ac3d300587b84f183dcd1c7f508cd, cited By (since 1996)15
18. Valkenborg, D., Mertens, I., Lemière, F., Witters, E., Burzykowski, T.: The isotopic distribution conundrum. Mass spectrometry reviews 31(1), 96–109 (2012), http://dx.doi.org/10.1002/mas.20339

Tables

Table 1. Basic Information on Stable Isotopes, as found in [15].

Element	Isotope	Extra Neutrons	Mass [Da]	Probability
Carbon	^{12}C	0	12	0.9893
	^{13}C	1	13.0033	0.0107
Hydrogen	^{1}H	0	1.0078	0.999885
	^{2}H	1	2.0141	0.000115
Nitrogen	^{14}N	0	14.0031	0.99632
	^{15}N	1	15.0001	0.00368
Oxygen	^{16}O	0	15.9949	0.99757
	^{17}O	1	16.9991	0.00038
	^{18}O	2	17.9992	0.00205
Sulfur	^{32}S	0	31.9721	0.9493
	^{33}S	1	32.9714	0.0076
	^{34}S	2	33.9679	0.0429
	^{36}S	4	35.9671	0.0002

Appendix

Proof of Lemma 3

We want to prove that if $\mu^{[n]} \rightharpoonup \mu$ and $\mu^{[n]}(A), \mu(A) > 0$, then also $\mu_A^{[n]} \rightharpoonup \mu_A$. We do this under the assumption that both $\mu^{[n]}$ and μ are discrete measures on probability space E.

By the *Portmanteau Lemma*, see [9], $\mu^{[n]} \rightharpoonup \mu$ implies that for any set A with boundry ∂A subject to $\mu(\partial A) = 0$, one should observe

$$\lim_{n\to\infty} \mu^{[n]}(A) = \mu(A). \tag{14}$$

The notion of boundry requires the notion of topology: thus, we decide on the discrete topology, which is natural in this context [6]. In this topology however, $\partial A = \emptyset$, for it is a set theoretical difference of the closure and the interior, both of which are equal to A. Hence, $\mu(\partial A) = 0$. Thus, (14) always holds.

Ex definitione, $\mu^{[n]} \rightharpoonup \mu$ means, that for any bounded function $f : E \to \mathbb{R}$ one observes

$$\int f \mathrm{d}\mu^{[n]} \xrightarrow[n\to\infty]{} \int f \mathrm{d}\mu\,. \tag{15}$$

A simple calculation using both (14) and (15) completes the proof:

$$\int f \mathrm{d}\mu_A^{[n]} = \frac{\int f \mathrm{d}\mu^{[n]}}{\mu^{[n]}(A)} \xrightarrow[n\to\infty]{} \frac{\int f \mathrm{d}\mu}{\mu(A)} = \int f \mathrm{d}\mu\,.$$

[6] For appropriate topological notions consult [5].