Rozkłady zagregowanych wariantów izotopowych

Piotr Dittwald

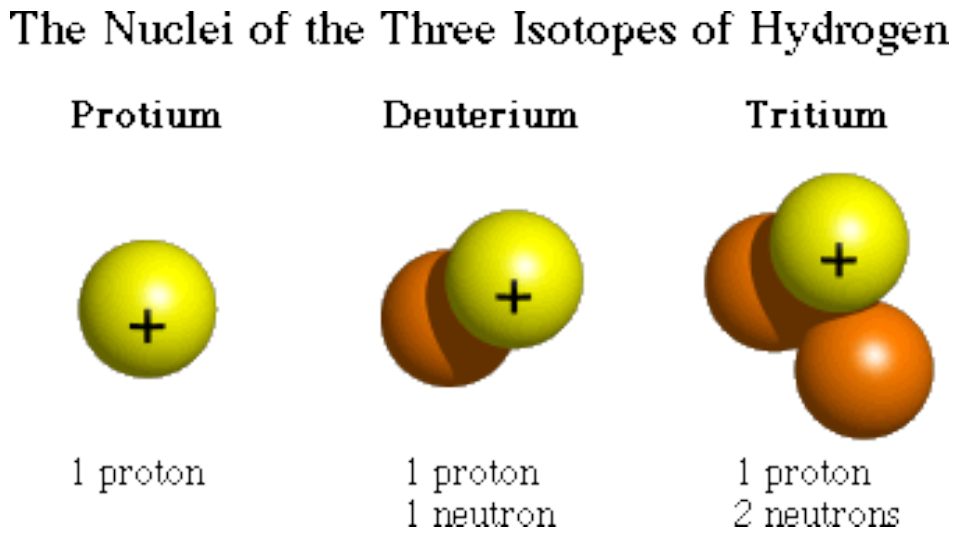
Uniwersytet Warszawski

9 I 2014

Przypomnienie: podstawowe definicje

Izotopy

warianty tego samego pierwiastka różniące się liczbą neutronów



source: http://education.jlab.org

Table 1: List of stable isotopes for carbon, hydrogen, nitrogen, oxygen, and sulphur.

Isotope	Mass (ma/u)	Abundance (%)	Isotope	Mass (ma/u)	Abundance (%)
^{12}C	12.0000000000	98.93	^{16}O	15.9949146	99.757
^{13}C	13.0033548378	1.07	^{17}O	16.9991312	0.038
^{1}H	1.0078250321	99.9885	^{18}O	17.9991603	0.205
^{2}H	2.0141017780	0.0115	^{32}S	31.97207070	94.93
^{14}N	14.0030740052	99.632	^{33}S	32.97145843	0.76
^{15}N	15.0001088984	0.368	^{34}S	33.96786665	4.29
			^{36}S	35.96708062	0.02

źródło: Cleasen et al., 2012, JASMS

Wariant monoizotopowy (bez dodatkowych neutronów)

Rozpatrujemy białko $C_v H_w N_x O_y S_z$. Łatwo jest policzyć prawdopodobieństwo oraz masę jego wariantu izotopowego:

$$P = P^v_{C_{12}} \times P^w_{H_1} \times P^x_{N_{14}} \times P^y_{O_{16}} \times P^z_{S_{32}}$$

$$M = vm_{C_{12}} + wm_{H_1} + xm_{N_{14}} + ym_{O_{16}} + zm_{S_{32}}$$

Jak policzyć pozostałe kombinacje?

Przykład: 3 atomy węgla (C)

$$(C_{12} + C_{13})^3 = C_{12}C_{12}C_{12} + C_{12}C_{12}C_{13} + C_{12}C_{13}C_{12} + C_{13}C_{12}C_{12} + C_{12}C_{13}C_{13} + C_{13}C_{13}C_{12} + C_{13}C_{12}C_{13} + C_{13}C_{13}C_{13}$$

Jak policzyć pozostałe kombinacje?

Przykład: 3 atomy węgla (C)

$$(C_{12} + C_{13})^3 = C_{12}C_{12}C_{12} + C_{12}C_{12}C_{13} + C_{12}C_{13}C_{12} + C_{13}C_{12}C_{12} + C_{12}C_{13}C_{13} + C_{13}C_{13}C_{12} + C_{13}C_{12}C_{13} + C_{13}C_{13}C_{13}$$

Ignorujemy kolejność

$$(C_{12} + C_{13})^3 = (C_{12})^3 + (C_{13})^3 + 3(C_{12})^2 C_{13} + 3C_{12}(C_{13})^3$$

Jak policzyć pozostałe kombinacje?

Przykład: 3 atomy węgla (C)

$$(C_{12} + C_{13})^3 = C_{12}C_{12}C_{12} + C_{12}C_{12}C_{13} + C_{12}C_{13}C_{12} + C_{13}C_{12}C_{12} + C_{12}C_{13}C_{13} + C_{13}C_{13}C_{12} + C_{13}C_{12}C_{13} + C_{13}C_{13}C_{13}$$

Ignorujemy kolejność

$$(C_{12} + C_{13})^3 = (C_{12})^3 + (C_{13})^3 + 3(C_{12})^2 C_{13} + 3C_{12}(C_{13})^3$$

To samo podejście dla ozonu (O_3)

$$(O_{16} + O_{17} + O_{18})^3 = (O_{16})^3 + (O_{17})^3 + (O_{18})^3 + 3(O_{16})^2 O_{17} + 3O_{16}(O_{17})^2 + 3(O_{16})^2 O_{18} + 3(O_{17})^2 O_{18} + 3O_{16}(O_{18})^2 + 3O_{17}(O_{18})^2 + 6O_{16}O_{17}O_{18}$$

Jak policzyć pozostałe kombinacje?

Przykład: 3 atomy węgla (C)

$$(C_{12} + C_{13})^3 = C_{12}C_{12}C_{12} + C_{12}C_{12}C_{13} + C_{12}C_{13}C_{12} + C_{13}C_{12}C_{12} + C_{12}C_{13}C_{13} + C_{13}C_{13}C_{12} + C_{13}C_{12}C_{13} + C_{13}C_{13}C_{13}$$

Ignorujemy kolejność

$$(C_{12} + C_{13})^3 = (C_{12})^3 + (C_{13})^3 + 3(C_{12})^2 C_{13} + 3C_{12}(C_{13})^3$$

To samo podejście dla ozonu (O_3)

$$(O_{16} + O_{17} + O_{18})^3 = (O_{16})^3 + (O_{17})^3 + (O_{18})^3 + 3(O_{16})^2 O_{17} + 3O_{16}(O_{17})^2 + 3(O_{16})^2 O_{18} + 3(O_{17})^2 O_{18} + 3O_{16}(O_{18})^2 + 3O_{17}(O_{18})^2 + 6O_{16}O_{17}O_{18}$$

Czy można to jeszcze uprościć?

Rozwinięcie wielomianowe

Dla białka o wzorze $C_v H_w N_x O_y S_z$ wprowadzamy wielomian z indykatorem I reprezentującym liczbę dodatkowych neutronów

$$\begin{aligned} Q(I; v, w, x, y, z) = &\left(P_{C_{12}} I^0 + P_{C_{13}} I^1\right)^v \times \\ &\left(P_{H_1} I^0 + P_{H_2} I^1\right)^w \times \\ &\left(P_{N_{14}} I^0 + P_{N_{15}} I^1\right)^x \times \\ &\left(P_{O_{16}} I^0 + P_{O_{17}} I^1 + P_{O_{18}} I^2\right)^y \times \\ &\left(P_{S_{32}} I^0 + P_{S_{33}} I^1 + P_{S_{34}} I^2 + P_{S_{36}} I^4\right)^z \\ = &\{Q_C(I)\}^v \times \{Q_H(I)\}^w \times \{Q_N(I)\}^x \times \{Q_O(I)\}^y \times \{Q_S(I)\}^z \end{aligned}$$

gdzie $P_{C_{12}}$, $P_{C_{13}}$, ..., $P_{S_{36}}$ odpowiadają prawdopodobnieństwom występowania odpowiednich izotopów.

Zagregowane warianty izotopowe

Z drugiej strony ten wielomian w standardowym zapisie

$$Q(l; v, w, x, y, z) = \sum_{j=0}^{n} q_j l^j$$

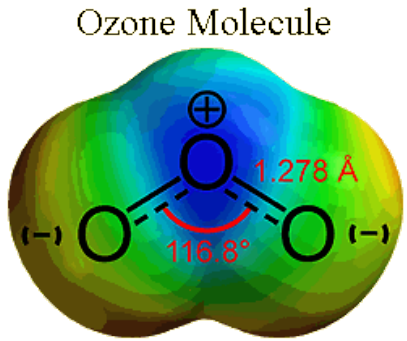
ma współczynniki $q_0, q_1, q_2, \ldots$ odpowiadające prawdopodobieństwom występowania wariantów izotopowych z, odpowienio, $0, 1, 2, \ldots$ dodatkowymi neutronami.

Przykład: ozon (O_3)

$$Q(I;0,0,0,3,0) = \left(P_{O_{16}} I^0 + P_{O_{17}} I^1 + P_{O_{18}} I^2\right)^3 = \sum_{j=0}^{6} q_j I^j$$

gdzie współczynniki $q_0, \ldots, q_6$ otrzymujemy jak następuje:

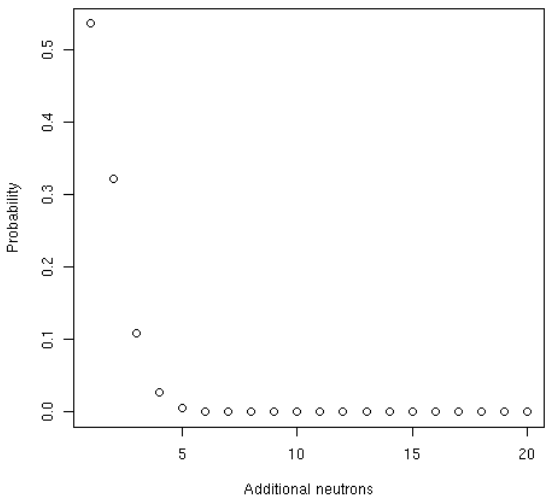
$$\begin{aligned}
q_0 &= P_{O_{16}}^3, \\
q_1 &= 3P_{O_{16}}^2 P_{O_{17}}, \\
q_2 &= 3P_{O_{16}}^2 P_{O_{18}} + 3P_{O_{16}} P_{O_{17}}^2, \\
q_3 &= P_{O_{17}}^3 + 6P_{O_{16}} P_{O_{17}} P_{O_{18}}, \\
q_4 &= 3P_{O_{17}}^2 P_{O_{18}} + 3P_{O_{16}} P_{O_{18}}^2, \\
q_5 &= 3P_{O_{17}} P_{O_{18}}^2, \\
q_6 &= P_{O_{18}}^3
\end{aligned}$$



source:
http://www.
ozonegreenheartland.com

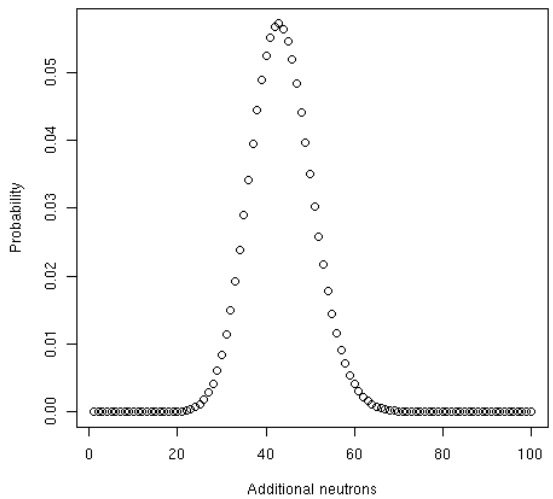
Przykłady: wybrane biomolekuły

$C_{50}H_{71}N_{13}O_{12}$



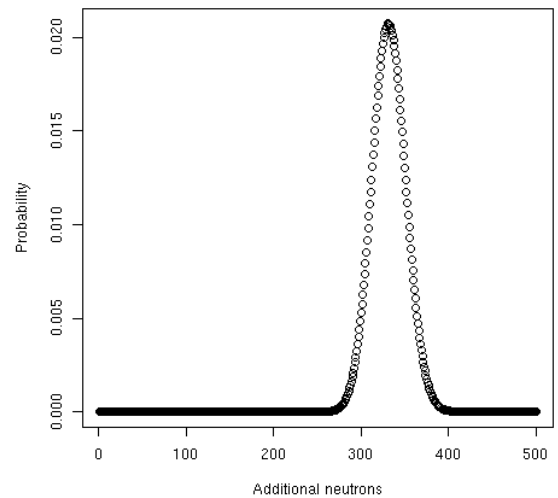
Angiotensin II

$C_{2934}H_{4615}N_{781}O_{897}S_{39}$



Bovine serum albumin

$C_{23832}H_{37816}N_{6528}O_{7031}S_{170}$



Human dynein heavy chain

Zagregowane warianty izotopowe

Problem

Obliczyć efektywnie $q_0, q_1, q_2, \ldots$

Zagregowane warianty izotopowe

Problem

Obliczyć efektywnie $q_0, q_1, q_2, \ldots$

Odpowiedź

Np. za pomocą szybkiej transformaty Fouriera (Fast Fourier Transform; FFT) do mnożenia wielomianów

Zagregowane warianty izotopowe

Problem

Obliczyć efektywnie $q_0, q_1, q_2, \ldots$

Odpowiedź

Np. za pomocą szybkiej transformaty Fouriera (Fast Fourier Transform; FFT) do mnożenia wielomianów

zadanie domowe

Zagregowane warianty izotopowe

Problem

Obliczyć efektywnie $q_0, q_1, q_2, \ldots$

Odpowiedź

Np. za pomocą szybkiej transformaty Fouriera (Fast Fourier Transform; FFT) do mnożenia wielomianów

zadanie domowe

Alternatywne podejście

Podejście algebraiczne - algorytm BRAIN (Claesen et al., 2012)

Średnia masa dla wariantów izotopowych

$$\bar{m}_j = \frac{\sum_k m_{jk} p_{jk}}{\sum_k p_{jk}},$$

gdzie $\bar{m}_j$ to średnia masa j-tego zagregowanego wariantu, a p_{jk} oraz m_{jk} znaczają prawdopodobieństwo i masę k-tego wariantu izotopowego wchodzącego w skład j-tego wariantu zagregowanego.

Nowa funkcja tworząca

$$U(I; v, w, x, y, z) = \sum_j \left(\sum_k m_{jk} p_{jk} \right) I^j \equiv \sum_j q_j^\star I^j$$

Nowa funkcja tworząca

$$U(I; v, w, x, y, z) = \sum_j \left(\sum_k m_{jk} p_{jk} \right) I^j \equiv \sum_j q_j^\star I^j$$

Teraz potrzebujemy obliczyć $q_j^\star$

$$Q^*(I, K; v, w, x, y, z) =$$
$$\left(P_{C_{12}} K^{M_{C_{12}}} I^0 + P_{C_{13}} K^{M_{C_{13}}} I^1\right)^v \times \left(P_{H_1} K^{M_{H_1}} I^0 + P_{H_2} K^{M_{H_2}} I^1\right)^w \times$$
$$\left(P_{N_{14}} K^{M_{N_{14}}} I^0 + P_{N_{15}} K^{M_{N_{15}}} I^1\right)^x \times$$
$$\left(P_{O_{16}} K^{M_{O_{16}}} I^0 + P_{O_{17}} K^{M_{O_{17}}} I^1 + P_{O_{18}} K^{M_{O_{18}}} I^2\right)^y \times$$
$$\left(P_{S_{32}} K^{M_{S_{32}}} I^0 + P_{S_{33}} K^{M_{S_{33}}} I^1 + P_{S_{34}} K^{M_{S_{34}}} I^2 + P_{S_{36}} K^{M_{S_{36}}} I^4\right)^z ,$$

gdzie $M_{C_{12}}$, $M_{C_{13}}$, ..., $M_{S_{36}}$ to masy izotopów węgla, wodoru, azotu, tlenu oraz siarki

Wielomian $Q^*(I, K; v, w, x, y, z)$ możemy wyrazić w następujący sposób:

$$Q^*(I, K; v, w, x, y, z) \equiv \sum_j \left(\sum_k p_{jk} K^{m_{jk}} \right) I^j$$

Wielomian $Q^*(I, K; v, w, x, y, z)$ możemy wyrazić w następujący sposób:

$$Q^*(I, K; v, w, x, y, z) \equiv \sum_j \left(\sum_k p_{jk} K^{m_{jk}} \right) I^j$$

Następny krok

Używamy $Q^*(I, K; v, w, x, y, z)$ aby uzyskać wielomian $U(I; v, w, x, y, z)$.

$$\frac{\partial}{\partial K} Q^*(I, K; v, w, x, y, z) = \sum_j \left(\sum_k m_{jk} p_{jk} K^{m_{jk}-1} \right) I^j$$

Podstawiając $K = 1$ otrzymujemy:

$$
\begin{aligned}
U(I; v, w, x, y, z) = \\
vQ(I; v-1, w, x, y, z)\left(P_{C_{12}} M_{C_{12}} + P_{C_{13}} M_{C_{13}} I^1\right) \\
+wQ(I; v, w-1, x, y, z)\left(P_{H_1} M_{H_1} + P_{H_2} M_{H_2} I^1\right) \\
+xQ(I; v, w, x-1, y, z)\left(P_{N_{14}} M_{N_{14}} + P_{N_{15}} M_{N_{15}} I^1\right) \\
+yQ(I; v, w, x, y-1, z)\left(P_{O_{16}} M_{O_{16}} + P_{O_{17}} M_{O_{17}} I^1 + P_{O_{18}} M_{O_{18}} I^2\right) \\
+zQ(I; v, w, x, y, z-1) \times \\
\left(P_{S_{32}} M_{S_{32}} + P_{S_{33}} M_{S_{33}} I^1 + P_{S_{34}} M_{S_{34}} I^2 + P_{S_{36}} M_{S_{36}} I^4\right) .
\end{aligned}
$$

Porównanie z innymi algorytmami

Table 2: List of selected biomolecules.

No.	Common Name	Molecular Formula	Mass (Da)	
			Monoisotopic	Average
(1)	Angiotensin II	$C_{50}H_{71}N_{13}O_{12}$	1045.534515	1046.181107
(2)	Bovine insulin	$C_{254}H_{377}N_{65}O_{75}S_6$	5729.600867	5733.510759
(3)	Human insulin	$C_{520}H_{817}N_{139}O_{147}S_8$	11616.849350	11624.448751
(4)	Human myoglobin	$C_{744}H_{1224}N_{210}O_{222}S_5$	16812.954775	16823.321352
(5)	Human intrinsic factor	$C_{2023}H_{3208}N_{524}O_{619}S_{20}$	45387.007033	45415.679370
(6)	Bovine serum albumin	$C_{2934}H_{4615}N_{781}O_{897}S_{39}$	66389.862474	66432.455561
(7)	Human Na/K ATPase Renal isoform, subunit	$C_{5047}H_{8014}N_{1338}O_{1495}S_{48}$	112823.879546	112895.125932
(8)	Human ATP binding cassette protein	$C_{8574}H_{13378}N_{2092}O_{2392}S_{77}$	186386.799265	186506.052594
(9)	Human intrinsic factor -hydroxocobalamin receptor	$C_{17600}H_{26474}N_{4752}O_{5486}S_{197}$	398470.366994	398722.972484
(10)	Human dynein heavy chain	$C_{23832}H_{37816}N_{6528}O_{7031}S_{170}$	533403.475090	533735.214651

źródło: Cleasen et al., 2012, JASMS

Wysokoprzepustowe obliczenia zagregowanych wariantów izotopowych

- ludzkie białka z bazy Uniprot,

Wysokoprzepustowe obliczenia zagregowanych wariantów izotopowych

- ludzkie białka z bazy Uniprot,
- przetwarzanie danych:

Wysokoprzepustowe obliczenia zagregowanych wariantów izotopowych

- ludzkie białka z bazy Uniprot,
- przetwarzanie danych:
  1. sekwencje $\longrightarrow C_v H_w N_x O_y S_z$,

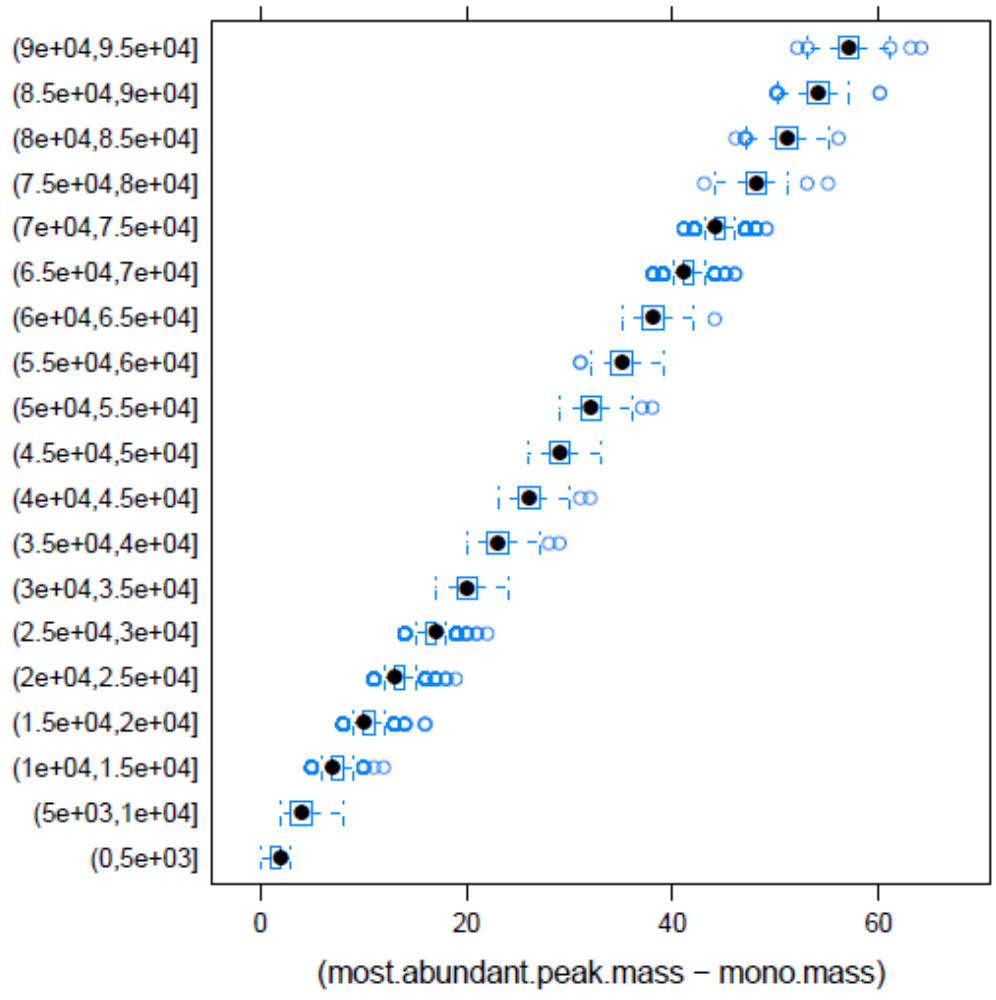
Wysokoprzepustowe obliczenia zagregowanych wariantów izotopowych

- ludzkie białka z bazy Uniprot,
- przetwarzanie danych:
  1. sekwencje $\longrightarrow C_v H_w N_x O_y S_z$,
  2. zagregowane warianty izotopowe,

Wysokoprzepustowe obliczenia zagregowanych wariantów izotopowych

- ludzkie białka z bazy Uniprot,
- przetwarzanie danych:
  1. sekwencje $\longrightarrow C_v H_w N_x O_y S_z$,
  2. zagregowane warianty izotopowe,
- w dalszej analizie rozważamy tylko białka o masie monoizotopowej poniżej 10^5 Da,

Wyniki



Oś y: Masa monoizotopowa

Oś x: różnica między masą najwyższego zagregowanego wariantu a masą monoizotopową

Model liniowy

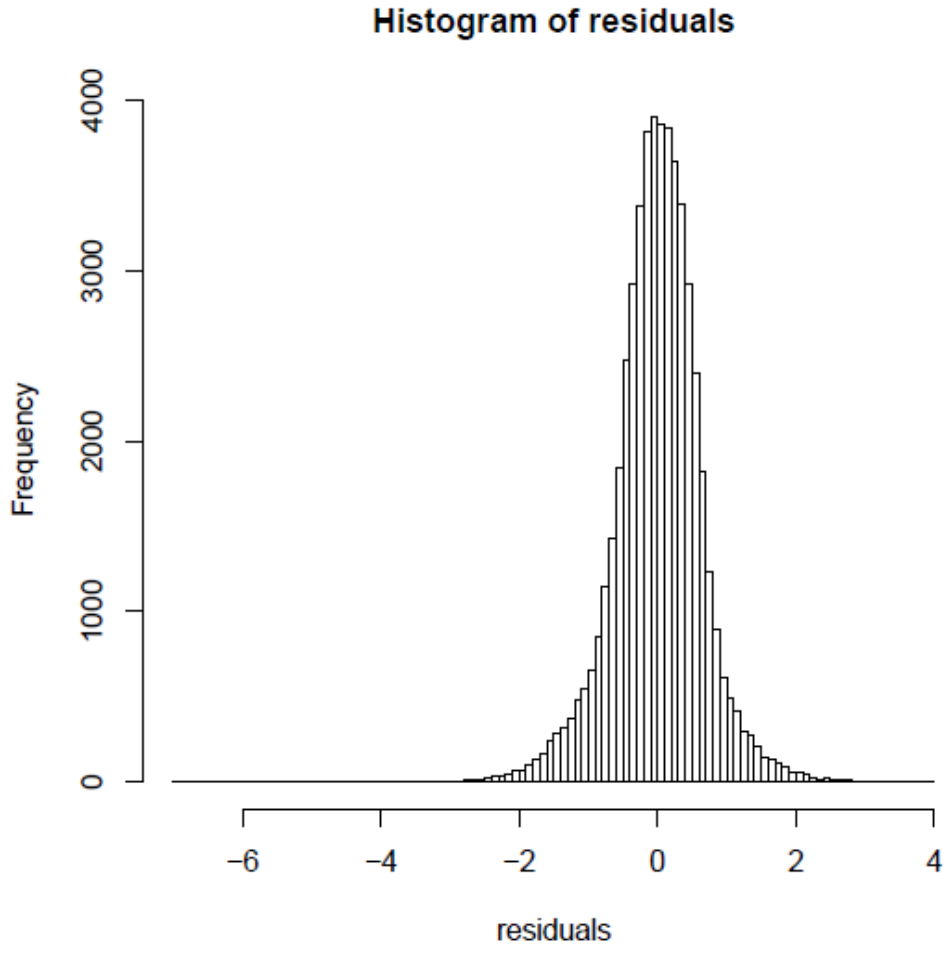
```
Call:
lm(formula = mono.mass ~ most.abundant.peak.mass, data = uniprot.10to5)

Residuals:
     Min       1Q   Median       3Q      Max
-6.93273 -0.34177  0.02066  0.37785  3.90323

Coefficients:
                          Estimate Std. Error   t value Pr(>|t|)
(Intercept)              4.820e-01  4.928e-03 9.781e+01   <2e-16 ***
most.abundant.peak.mass 9.994e-01  1.220e-07 8.193e+06   <2e-16 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 0.6557 on 52628 degrees of freedom
Multiple R-squared:     1,        Adjusted R-squared:     1
F-statistic: 6.712e+13 on 1 and 52628 DF,  p-value: < 2.2e-16
```

Residua modelu liniowego



Dodatek: średnia masa dla całej cząsteczki

- Najprościej obliczyć za pomocą formuły zamkniętej:

$$\bar{m} = v\bar{m}_C + w\bar{m}_H + x\bar{m}_N + y\bar{m}_O + z\bar{m}_S$$

  gdzie $\bar{m}_C, \bar{m}_H, \bar{m}_N, \bar{m}_O, \bar{m}_S$ to średnie masy poszczególnych atomów, np. $\bar{m}_C = p_{C_{12}} m_{C_{12}} + p_{C_{13}} m_{C_{13}}$
- Alternatywnie można też zsumować wszystkie (ew. pomijając warianty o dostatecznie małym prawdopodobieństwie) iloczyny prawdopodobieństw i poszczególnych mas dla wariantów izotopowych

Współpraca

Anna Gambin[a], Jürgen Claesen [b], Tomasz Burzykowski [b], Dirk Valkenborg [b,c,d]
a : University of Warsaw, Poland
b : Hasselt University, Belgium
c : Flemish Institute for Technological Research (VITO), Belgium
d : CfP-CeProMa, University of Antwerp, Belgium

Bibliografia

- Claesen, J, Dittwald, P, Burzykowski, T, Valkenborg, D (2012). An efficient method to calculate the aggregated isotopic distribution and exact center-masses. J. Am. Soc. Mass Spectrom., 23, 4:753-63.
- Valkenborg, D, Mertens, I, Lemière, F, Witters, E, Burzykowski, T (2012). The isotopic distribution conundrum. Mass Spectrom Rev, 31, 1:96-109.
- Dittwald, P, Claesen, J, Burzykowski, T, Valkenborg, D, Gambin, A (2013). BRAIN: A Universal Tool for High-Throughput Calculations of the Isotopic Distribution for Mass Spectrometry. Anal. Chem., 85, 4:1991-4.