

X_T and \hat{Y}_T should have similar histograms.

$$L(Y_T | X_T)$$

X_T - true

Y_T - observed

$$|X_T| = \text{sum} \left(\text{values}[\text{keys}] : \text{keys} \right)$$

$$X_T = \left\{ \begin{array}{l} \text{Count}[\text{observed}], \emptyset[\text{unobserved}] \end{array} \right\}$$

$$Y_T = \left\{ \begin{array}{ll} AB \rightarrow 1 & \frac{1}{6} \\ A \rightarrow 3 & \frac{1}{2} \\ C \rightarrow 2 & \frac{1}{3} \end{array} \right.$$

6

$$X_T = \left\{ \begin{array}{l} AB \rightarrow 1 \\ A \rightarrow 2 \end{array} \right.$$

Possibilities:

A, AB, ABC, B, BC, C

Hamiltonian:

$$\propto \exp(-)$$

$$M(|X_T|)$$

Blaise: $\begin{pmatrix} A \\ \text{Observed} \\ A_1, \dots, A_n \end{pmatrix}$ Each peak independently changed

$$\prod_{i=1}^n \left(\frac{X_i + \epsilon}{\sum X_i + n\epsilon} \right)^{A_i} \sim \sum_{i=1}^n A_i \left[\log(X_i + \epsilon) - \log\left(\sum_{i=1}^n X_i + n\epsilon\right) \right] \sim$$

$$\sim \sum_{i=1}^n A_i \log(X_i + \epsilon) = \sum_{\substack{X_i > 0 \\ A_i > 0}} A_i \log(X_i + \epsilon) + \underbrace{\left(\sum_{X_i = 0} A_i \right)}_{\substack{\text{potential for } -\infty \\ \text{put exception here}}} \log(\epsilon)$$

Other possibilities

$$-\sum_i \left| \frac{A_i}{A} - \frac{X_i}{X} \right| = -\|P_A - P_X\|_{L_1}$$

$$-\|P_A - P_X\|_2^2$$

Total N° of possibilities

Is it needed? No

3x Counter

$$\|P_A - P_B\| = \sum_{\substack{X_i = 0 \\ A_i > 0}} P_{A_i} + \sum_{\substack{A_i > 0 \\ X_i > 0}} |P_{A_i} - P_{X_i}| + \sum_{\substack{X_i > 0 \\ A_i = 0}} P_{X_i}$$

(FALSE TRUE) (F T) (T F)

Problem:

? A_i are censored data really.
 \hookrightarrow We observe only A_i s.t. $A_i \text{ change} > 0$

$$\sum_{A_i > 0} A_i \log(X_i + \epsilon)$$

One simple calculation

$$\frac{e^{-\|P_A - P_X\|}}{\sum_S e^{-\|P_A - P_B\|}}$$

S : obtainable histograms with a given n° of molecules (diff. granularity)