Probabilistic model of random co-colocalization in a cell

April 16, 2025

1 Aim

For biological studies purpose we image a cell in 3D using fluorescent microscopy and we are interested in quantifying the interactions between different types of single molecules. From those images we have no means of quantifying directly the interaction between single molecules, instead we want to see how frequently they localize together in space. Our goal is to build a general model that will modelise the likelyness that a single molecule from distribution i localize together or **co-localize** with a single molecule from distribution j in a cell of volume V **assuming single molecules take random postions within the cells independently of one another**.

Our images are produced through a sequential fish microscope in fixed cells that divides, according to its resolution cells in a set of v voxels localized with their set of coordinates *(z,y,x)* in the volume V.

2 Co-localization events

2.1 Self co-localization event

We note *C(i,i)* a **self co-localization** event between two single molecules of distribution i occupy the same voxel.

2.2 Co-localization event

We note *C(i,j)* a **co-localization** event between two single molecules of distribution i and *(*j) when a single molecule from distribution i occupy the same voxel as a single molecule from distribution j.

3 Probabilistic model

3.1 Presentation

The process of assigning a position to single molecules amongst the v possible positions can be modeled as a probabilistic game where positions are uniquely numbered balls placed in an urn. Assigning coordinates to a single molecule is drawing a ball in the urn, before each drawn balls are replaced in the urn.

3.2 Unique distribution and self co-localization

To begin, let us consider a system containing a unique distribution $\mathbf{I}$ of single molecules randomly placed amongst the v positions of the cell. The probability that a specific single molecule i is found at position m is the probability to draw the ball numbered m:

$$p_i(X = m) = \frac{1}{v} \qquad \forall i \in I; \forall m \in V$$

Then the probability to not draw the specific location m is :

$$p_i(\overline{m}) = \frac{1}{v} \qquad \forall i \in I; \forall m \in V$$

After k draws, the probability to never draw a the specific location m follows a bionomial law and is :

$$p_k(X_m = 0) = (1 - \frac{1}{v})^k \qquad k \in \mathbb{N}$$

On the contrary the probability that the location m was drawn at least once is :

$$p_k(X_m \geq 1) = 1 - (1 - \frac{1}{v})^k \qquad \textbf{(i)}$$

To study self co-localization we are interested to know how many different positions have been drawn in k-trials. To do so let us define the observation variable ϵ.

$$\forall m \in V, I(m) = \begin{cases} 0 \text{ if } X_m = 0 \\ 1 \text{ if } X_m \geq 1 \end{cases}$$

The expected number of **different** positions N_{pos}^{unique} drawn is then the expectancy of ϵ after k draws.

$$N_{uniquepos} = \sum_{m \in V} p(X_m = 0).0 + p(X_m \geq 1).1$$

$$N_{uniquepos} = v(1 - (1 - \frac{1}{v})^k) \qquad \textbf{(ii)}$$

The expected number of different positions drawn is, in other words, the expected number of draws **that discovered a new position**. To adress self co-localization probability we are interested to know the number of draws that **didn't** discover a new position, in other words, duplicates draws or **self co-localization events**. To do so, we remove from the total number of picks (k) the number of picks that discover a new position.

$$N_{selfcolocalization} = k - v(1 - (1 - \frac{1}{v})^k)$$

$$p_{selfcolocalization} = 1 - \frac{v}{k}(1 - (1 - \frac{1}{v})^k)$$

Notes:

- This is true for any distribution I of single molecule i of abundancy k_i (i.e. number of single molecule).

- $N_{selfcolocalization}$ is the expectancy E of a bionomial law of sucess probability $p = p_{selfcolocalization}$

- Standard deviation of the expected count of self co-localization can be computed as $np(1 - p)$ (bionomial law)

3.3 2 distributions and co-localization probability

We now consider positions assignement of a second distribution I of abundancy k_i after the assignement of a first distribution J of abundancy k_j with $I \neq J$. The number of unique positions J occupy can be estimated with **(ii)**.

$$N_{uniquepos}^{j} = v(1 - (1 - \frac{1}{v})^{k_j})$$

A colocalisation event is drawing during assignement of I a position that was already drawn when assigning positions to J. To understand this in our probabilistic game, let us consider again a pool of v uniquely numbered balls where all balls that were drawn while assigning postions to J distribution have been colored in **red**. Again all balls are replaced in the pool after each draw. Then the co-localization probability $p(C(i, j))$ is the probability to draw a red ball.

3.3.1 Probability of co-localization

Drawing ball in the urn is a sequence of independent and uniformly random events. The probability of picking one of the $N_{uniquepos}^{j}$ red ball amongst the v balls is :

$$p(X \in \mathbb{J}) = \frac{N_{uniquepos}^{j}}{v}$$

where $\mathbb{J}$ is the set of positions drawn for the J distribution. Again co-localization events follow a bionomial process of sucess probability $p = p(X \in \mathbb{J})$. This co-localization probability can also be interpreted as the co-localization rate of molecules i with molecules j.

We can deduce the expectancy and standard deviation of co-localization events :

$$E(C(i,i)) = k_i \frac{N^j_{uniquepos}}{v} \qquad \textbf{(iii)}$$

$$std(C(i,j)) = k_i \frac{N^j_{uniquepos}}{v}(1 - \frac{N^j_{uniquepos}}{v})$$

Notes:

- $E(C(i,j) \neq C(j,i))$

3.3.2   Number of unique positions occupied by a pair (i,j)

In section 3.1 we deduced the probability that a specific location m to be drawn at least once in k draws as :

$$p_k(X_m \geq 1) = 1 - (1 - \frac{1}{v})^k \qquad \textbf{(i)}$$

To know how many red balls have been drawn at least once (**i.e. number of unique pair (i,j)**) let us use again the observation variable I.

$$N^{(i,j)}_{uniquepos} = \sum_{m \in \mathbb{J}} p(X_m \geq 1).1 + 0. \ ...$$

$$N^{(i,j)}_{uniquepos} = N^j_{uniquepos}(1 - (1 - \frac{1}{v})^{k_i}) \qquad \textbf{(iv)}$$

or :

$$N^{(i,j)}_{uniquepos} = v(1 - (1 - \frac{1}{v})^{k_j})(1 - (1 - \frac{1}{v})^{k_i})$$

Notes:

- Though $E(C(i,j)) \neq E(C(j,i))$, explicit writting in last equation shows that $N^{(i,j)}_{uniquepos} = N^{(j,i)}_{uniquepos}$.

- Last equation also seems to show it will be easy to generalise the count of unique position occupied by any combination of molecules c of dimension n as $N^c_{uniquepos} = v \prod_{i \in c}(1 - (1 - \frac{1}{v})^{k_i})$.

3.3.3 Standard deviation of unique positions number occupied by a pair (i,j)

The number of different positions occupied by pairs (i,j), $N^{(i,j)}_{uniquepos}$, is actually expectancy of the sum of **dependent** random variables $I(m) \forall m \in V$. Let us call $U(i,j)$ the random variable describing this.

$$U(i,j) = \sum_{m \in \mathbb{J}} I(m)$$

$$N^{(i,j)}_{uniquepos} = E(U(i,j))$$

We already determined the expectancy of the random variable U in above subsection. To fully define the probability density of U let us find its standard deviation. The sum of random variables follows the law of **total variance** :

$$Var(U(i,j)) = \sum_{m \in \mathbb{J}} Var(I_m) + 2. \sum_{(l,m) \in \mathbb{J}^{\neq}} Cov(I_m, I_l)$$

To simply calculus, let us break it into 2 parts :

$$S = \sum_{m \in \mathbb{J}} Var(I_m)$$

$$C = \sum_{(l,m) \in \mathbb{J}^{\neq}} Cov(I_m, I_l)$$