Homework 5

Due 2/27 4pm

---

1. 3 Advantages of Counting mRNA Molecules to Measure Gene Expression:

- Discrete Numbers: We can quantify the 'exact' number of mRNA moleucules that contribute to protein expression. We can also compare this to the total number of mRNA molecules to get a better picture of how much stuff is expressing the protein.
- More stuff: We can account for a parts of the cell that are low in intensity but still contribute to the overall expression of the protein.
- We are not counting empty space where there are no mRNA molecules. The intensity recording method would count regions of the cell(that do not have any mRNA molecules) as a data point which could skew the results.

2. 3-5 Potential Sources of Error

- Is the field of view (FOV) representative of the entire cell? The location of where this image is taken with respect to the cell could be a major source of error; we could make a mistake by taking a picture of the same small region across multiple cells, and this would not accurately represent the entire cell.
- Is the labeling process accurate? Does it account for all the mRNA molcules that relate to protein expression, and can we be sure that it doens't also label other unrelated molecules which have nothing to do with protein expression? Also
- How do we know what counts as a single molecule in relation to a fluorescing spot? There are spots of different intensities and sizes, so it could be possible that a large bright spot is actually multiple mRNA molecules.
- How do we know that the mRNA molecules are not degrading over time? That is, there is a finite time for the mRNA to fluoresce before it degrades, and we could be missing data.

3. Most problematic: The labeling procees is the most problematic because there is so much biological complexity that we can't account for. Unless we have a perfect labeling process, there can be errors from not accounting for all the mRNA molecules and also accidentally labeling other molecules that are not related to protein expression.

4. Lets say we have N total number of particles in a volume V. The probability of finding n_o particles in a volume v_o can be given by the binomial distribution:

$$P(n_o) = \frac{N!}{n_o!(N-n_o)!} f^{n_o}(1-f)^{N-n_o}$$

where the frequency

$$f = \frac{\lambda}{N} = \frac{rT}{N}$$

is the ratio of the average number of particles λ to the total number N, and r is the rate of particles entering this volume over a time T. After some mathy stuff:

$$\begin{aligned} P(n_o) &= \frac{N(N-1)\dots(N-n_o+1)\cancel{(N-n_o)!}}{n_o!\cancel{(N-n_o)!}} \left(\frac{\lambda}{N}\right)^{n_o} \frac{\left(1-\frac{\lambda}{N}\right)^N}{\left(1-\frac{\lambda}{N}\right)^{n_o}} \\ &= \frac{N(N-1)\dots(N-n_o+1)}{N^{n_o}} \left(1-\frac{\lambda}{N}\right)^{-n_o} \frac{\lambda^{n_o}}{n_o!} \left(1-\frac{\lambda}{N}\right)^N \end{aligned}$$

and for large number of total particles $N \to \infty$, we have two terms that go to 1:

$$\begin{aligned}\frac{N(N-1)\dots(N-n_o+1)}{N^{n_o}} &= \frac{N}{N}\frac{N-1}{N}\cdots\frac{N-n_o+1}{N} \\ &= 1\left(1-\frac{1}{N}\right)\dots\left(1-\frac{n_o-1}{N}\right) \approx 1\end{aligned}$$

and

$$\left(1-\frac{\lambda}{N}\right)^{-n_o} \to 1$$

And from the limit definition of the exponential function:

$$\lim_{N\to\infty}\left(1+\frac{-\lambda}{N}\right)^{N} \to e^{-\lambda}$$

So we finally get

$$P(n_o) = \frac{\lambda^{n_o}}{n_o!}e^{-\lambda}$$

thus obeying Poisson statistics.

5. This 'anomaly' is perhaps due to the fact that the molecule counts are centered around the detected nucleus center. In Problem 4, we assumed that this test volume was a randomly chosen volume, and the researchers at Fancy University have chosen a test volume that is dependent on focusing around nucleus centers for each molecule count. This method would disregard volumes that are not related to a nucleus, so we have added a bias in our method of data analysis. If we were to exclude this automatic nucleus centering i.e. we define this $100\,\mu m^3$ cylinder randomly in our FOV, we may get something closer to Poissonian noise.