

# Homework 5

Jonathan Monroe

October 2018

## 1 Spot counting

A new method for estimating Hb expression is proposed. It consists of high-resolution images of individual mRNA molecules. Compare this to the pixel averaging done in homework 3 where stripes across the embryo were averaged to get a net profile.

### 1.1 Benefits of molecule counting

1. This counting method seems to have higher fidelity. Rather than enforcing pixels' "built in" scale (e.g. 0 to 255), counting is a more direct measurement which more directly maps to the science question "how much expression is here?"
2. Estimated error can be smaller for Poisson counting error than for Gaussian estimates.
3. This method does not include "dead space" of the pixel adding method. i.e. averaging over pixel values includes space between mRNA as background noise.

### 1.2 Pipeline drawbacks

Although the new counting methods benefit from at least the above benefits, the full pipeline is not without limitations.

1. The labelling process may not be effective. What fraction of mRNA molecules get tagged? With what efficiency are tagged molecules detected?
2. The imaging protocol could neglect significant slices of the embryo. Hopefully the researchers know which plane in the embryo contains the most representative mRNA density. E.g. the interior of the embryo could contain a higher density of mRNA than the exterior. Which of these is representative of the variation?
3. Background noise is still present and requires some threshold to differentiate from signal. The automated detection threshold could be set too low, thereby counting mRNA where there is none, or it could be set too high and miss expression.
4. Similarly, mRNA expression could be too localized as to be resolved with their detectors. e.g. the largest bright circle in the bottom right of Figure 1 is likely many mRNA patches.
5. Finally, the AP axis is harder to define because the "off" region outside a cell is harder to estimate when only using counting statistics. What threshold should one use to distinguish between a lull in the counting rate and a bona fide transition to cell exterior? Without these boundaries defined everywhere (especially on the posterior end) the cell length is hard to normalize.

### 1.3 Serious drawbacks

I believe the *staining* step is the most error prone. These measurements could be grossly misleading without independent calibration. One couldn't say what fraction of mRNA molecules are actually observed.

Besides accuracy, I'm also concerned about feasibility. If protein staining is hard, mRNA staining must be extra hard owing to size. My naive understanding of biology would say mRNA is harder to detect via imaging than proteins.

## 2 Expression Noise

The researchers note two remarkable facts:

1. Noise in the fully-on region is lower than in the intermediate region
2. Fluctuations in Kb and Hb are anti-correlated.

But are these observations real or merely artifacts?

### 2.1 Alternative explanations

1. A steep drop off is more sensitive to perturbations in position. Thus, variation in AP axis assignment (e.g. due to inaccurate edge detection) can lead to large variations within bins.
2. Poisson noise is correlated with its signal intensity. Therefore if the expressions are anti-correlated then their Poisson errors are also anti-correlated.

The first issue would need to be clarified first because the correlations are contingent on first having representative bins to correlate.

## 3 Noise limits

### 3.1 Deriving Poisson statistics

The particle number probabilities of a collection of freely diffusing non-interacting particles obey a Poisson distribution.

First, consider a small volume  $V_0$ . The probability of this small volume having  $m$  of  $N$  total particles requires  $V_0$  to have exactly  $m$  and every other volume outside  $V_0$  to contain the remaining  $N - m$  particles. Because the particles do not interact the presence of each particle is independent of other particles.

Define the probability of having  $m$  particles in a volume  $V_0$  has  $p = \frac{\bar{n}}{N}$ . The independent trials of placing  $N$  particles in boxes of size  $V_0$  follows a binomial distribution:

$$\begin{aligned}
 P(m) &= \lim_{N \rightarrow \infty} \binom{N}{m} p^m (1-p)^{N-m} \\
 &= \lim_{N \rightarrow \infty} \frac{N!}{(N-m)!m!} \left(\frac{\bar{n}}{N}\right)^m \left(1 - \frac{\bar{n}}{N}\right)^{N-m} \\
 &= \frac{\bar{n}^m}{m!} \lim_{N \rightarrow \infty} \frac{N(N-1)\dots(N-m+1)}{NN\dots N} \left(1 - \frac{\bar{n}}{N}\right)^N \left(1 - \frac{\bar{n}}{N}\right)^{-m} \\
 &= \frac{\bar{n}^m}{m!} \cdot 1 \cdot e^{-\bar{n}} \cdot 1
 \end{aligned}$$

which is a Poisson distribution.

Now, the observation of sub-Poisson noise could still be an artifact of the analysis. At the highest densities, individual molecules may be hard to count. The maximum count corresponds to 100 particles in  $100 \text{ um}^2$ . Yet the particles in Figure 1 are of order  $1 \text{ um}^2$ . These rates may be simply too high to count accurately with these methods.

If the counting reliably resolves the particles, the Poisson process may be an inappropriate model as the number of particles approaches the number of “bins” ( $\frac{V}{V_0} \sim N$ ).

However, all things considered, sub-Poisson noise could come from a variety of situations. Sub-Poisson statistics in quantum optics is known as anti-bunching. A state of a definite number of photons (a Fock state) impinging on a photodetector exhibits sub-Poissonian noise. Detecting this state implies that fewer-than-expected counts will occur in the interim.

Likewise, in this experiment sub-Poissonian statistics could arise from something like expression feedback in mRNA. Once mRNA has reached a particular density it could induce feedback in nearby cells to equilibrate with it.