

Exercise: What's my question?

Background: You are interested in using a GFP marker in *E. coli* to monitor stochastic changes in gene expression.

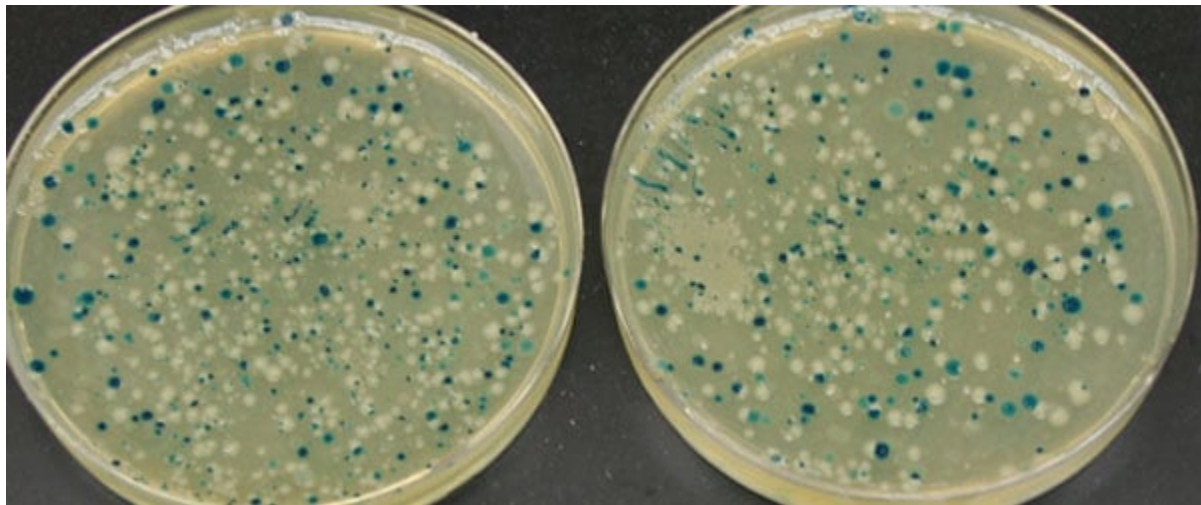
(Note: It is absolutely not necessary to read any background material for this exercise, but if you are curious about this topic, you could check out one of these reviews by Alex van Ouenarden:

[https://www.cell.com/cell/fulltext/S0092-8674\(08\)01243-9](https://www.cell.com/cell/fulltext/S0092-8674(08)01243-9) ;

<https://science.sciencemag.org/content/336/6078/183.abstract>)

Goal: You have just transformed a stock of *E. coli* with your GFP construct and plated them out on a large number of plates. Our exercise is to **describe one question that can be answered using each of the probability distributions below**.

The following picture illustrates how your plates might look (except it shows colonies expressing beta-galactosidase plated on IPTG):



Conditions:

- The probability of observing a cell with GFP may be anywhere between 0 and 1.
- You may consider just one plate, with a finite number of colonies on it, or a large number of plates (so that the total number of colonies is extremely large).
- You may ask questions about the chance of picking transformed vs. non-transformed cells in a random sample, the spatial distribution of transformed colonies on the plates, the temporal distribution of stochastic gene expression, or some other aspect of the experiment.

Distributions:

- Binomial (Bernoulli)
- Hypergeometric
- Negative binomial (Geometric)
- Poisson
- Exponential
- Normal

For each distribution, answer the following (no coding is required):

- What is the question you came up with for this scenario?
- More generally, what kind of problem is the function typically used for?
- How is the distribution parameterized?
- What do the PDF and CDF represent, in general terms?
- What do the PDF and CDF for this distribution look like, generally speaking?