AP3162: Stochastic models beyond gene regulations

Hyun Youk
Delft University of Technology
(Dated: March 6, 2020)

In this lecture, we discuss probabilistic models in biology, beyond those of cell division and stochastic gene-expression that we discussed in the previous lecture.

I. STATISTICAL QUANTITIES

The goal of this section is to review key statistical quantities.

A. Discrete probabilities

Definition of probability: Suppose that there are N possible outcomes in total ($N \geq 1$). Let X be the outcome value. Then, we define the **probability of getting an outcome of value of** X to be

$$P(X) = \frac{\text{number of ways of getting the value } X}{\text{total number of possible outcomes}} \tag{1}$$

Equivalently, we can also say

$$P(X) = \frac{\text{number of ways of getting the } \textbf{desired outcome-value } X}{\text{total number of possible outcomes}} \tag{2}$$

Note that

$$0 \leq P(X) \leq 1 \tag{3}$$

Probability of this AND that:

$$P((x,y)) = \frac{(\text{number of ways of getting x}) \cdot (\text{number of ways of getting y})}{\text{total number of outcomes}} \tag{4a}$$

$$= \frac{(\text{number of ways of getting x}) \cdot (\text{number of ways of getting y})}{(\text{total \# of outcomes for the } \textbf{first} \text{ experiment}) \cdot (\text{total \# of outcomes for } \textbf{second} \text{ experiment})} \tag{4b}$$

$$= \Big(\frac{\text{number of ways of getting x}}{\text{total \# of outcomes for the } \textbf{first} \text{ experiment}}\Big) \cdot \Big(\frac{\text{number of ways of getting y}}{\text{total \# of outcomes for the } \textbf{second} \text{ experiment}}\Big) \tag{4c}$$

$$= P(x) \cdot P(y) \tag{4d}$$

Motivated by this example, we can generalize as follows: First, we say "event" or "experiment" to refer to a process that yields an outcome value, such as throwing a dice. Another example of an event or an experiment is blindly picking a card from a deck of Poker cards. Moreover, suppose that we have two **events** that are **independent of each other** like in case of throwing of the dice twice - that is, the outcome of one event does not depend on the other events. If X is the outcome value of the first experiment and Y is the outcome value of the second experiment, then by above derivation, we can see that

$$P(X,Y) = P(X) \cdot P(Y) \tag{5}$$

We say that Eq. 5 is the **joint probability** for two independent events. In fact, we can generalize to M independent events. Using the derivation given in Eqs. 4a - 4d, it follows that if $X_1, X_2, ...X_M$ are the outcome values of each of the M independent experiments, then

$$P(X_1, X_2, ..., X_M) = P(X_1) \cdot P(X_2) \cdot ... \cdot P(X_M) \tag{6}$$

Probability of this OR that:
For two events, the probability of getting outcome values of **either** X **or** Y is

$$P(X \ \underline{\textbf{or}} \ Y) = P(X) + P(Y) \tag{7}$$

We can further to say that the probability of getting one of M desired outcome values (denoted X_1, X_2, ... X_M) out of a total of N possible outcomes in one experiment is

$$P(X_1 \ \underline{\textbf{or}} \ X_2 \ \underline{\textbf{or}} \ ... \ \underline{\textbf{or}} \ X_M) = P(X_1) + P(X_2) + ... + P(X_M) \tag{8}$$

Normalization condition: If there is a set of N possible, distinct outcome-values for an experiment, denoted by $\{x_1, x_2, ..., x_N\}$, then

$$P(x_1) + P(x_2) + ... + P(x_N) = 1 \tag{9}$$

We can compactly write Eq. 9 as

$$\sum_{i=1}^{N} P(x_i) = 1 \tag{10}$$

In summary, summing up the probabilities for each possible outcome should yield 1.

Definitions of statistical quantities:

Now that we know how to calculate probabilities of events, let's define five key **statistical quantities** - (1) **random variable**, (2) **average** (also called the mean or the **expectation value**), (3) **standard deviation**, (4) **variance**, and (5) **fractional error**.

(1) Random variable: A random variable X is a variable whose value is the outcome value of an experiment. Thus, there is a probability $P(x_i)$ of X taking on a particular value x_i. If there is a total of N possible outcome values (x_1, x_2,..., x_N), then X can take on any one of the N values. So we only know probabilistically, not exactly, the value of X. In the example of throwing a six-sided dice, the outcome-value is the random variable X, which can take on one of $\{1, 2, 3, 4, 5, 6\}$ as a value.

Note that a sum of two random variables is also a random variable. That is, if X and Y are random variables, then so is $X + Y$ because the summed value is also probabilistic - we cannot definitely predict the summed value before the experiment of measuring X and Y because we are unsure of the value of X and the value of Y before the experiment. For the same reason, a definite constant (e.g., 3) times a random variable is also a random variable (e.g., $3X$). Likewise, product of two random variables, XY, is also a random variable.

(2) Mean: The mean value (i.e., **expectation value**, **average**) of a random variable X is written as $< X >$ and defined as

$$< X >= x_1 P(x_1) + x_2 P(x_2) + ... + x_N P(x_N) \tag{11}$$

where $\{x_1, x_2, x_3, ..., x_N\}$ is the set of N possible, distinct outcome-values and $P(x_i)$ is the probability of getting the outcome-value x_i. Applying this definition to a throwing of the six-sided dice, the expectation value of the outcome is

$$< X > = 1P(1) + 2P(2) + ... + 6P(6) \tag{12a}$$
$$= \frac{1}{6} + \frac{2}{6} + ... + \frac{6}{6} \tag{12b}$$
$$= 3.5 \tag{12c}$$

which matches our expectation (thus the term, "expectation value"). We can write Eq. 11 more compactly as

$$< X >= \sum_{i=1}^{N} x_i P(x_i) \tag{13}$$

Eq. 13 is the definition of the mean-value of random variable X. There are two properties of the mean that we will use. The first deals with the mean of a new random variable that we form by multiplying a random variable X by a constant c. The resulting, new random-variable cX - note that this is random since we cannot definitely predict its value before doing an experiment - has the following mean:

$$< cX > = \sum_{i=1}^{N} c x_i P(x_i) \tag{14a}$$
$$= c \sum_{i=1}^{N} x_i P(x_i) \tag{14b}$$
$$= c < X > \tag{14c}$$

The second property of the mean deals with the mean of a new random-variable that we form by adding a constant c to a random variable X. The resulting, new random-variable $(c + X)$ is also a random variable for the same reason that cX is a random variable. Its mean is

$$< c + X > = \sum_{i=1}^{N} (c + x_i) P(x_i) \tag{15a}$$
$$= c \sum_{i=1}^{N} P(x_i) + \sum_{i=1}^{N} x_i P(x_i) \tag{15b}$$
$$= c \quad + < X > \tag{15c}$$

where we used the normalization condition (Eq. 10) in the last line.

(3) Standard deviation: The standard deviation σ (Greek letter "sigma") of a random variable X is defined as

$$\sigma = \sqrt{\left\langle (X - < X >)^2 \right\rangle} \tag{16}$$

The standard deviation (Eq. 16) quantifies the expected deviation of a random variable's value from the mean. You might wonder why the square and the square root are in Eq. 16. You might say that a more natural way measuring the expected (mean) deviation of a random variable's value from its mean would be

$$< X - < X >> \qquad \text{(why not this as the definition of } \sigma\text{?)} \tag{17}$$

But this would not work. In fact, for *any* random variable, we have

$$< X - < X >>= 0 \tag{18}$$

and thus this $< X - < X >>$ is not an informative quantity for what we want. To see why Eq. 18 is true, note that since $< X >$ is a constant, Eq. 15c tells us that

$$< X - < X >>=< X > - < X >= 0 \tag{19}$$

Eq. 19 also reveals a deeper meaning of the expectation value $< X >$: The mean $< X >$ is the value for which the random variable X fluctuates, on average, just as many times above $< X >$ (i.e., $(X - < X >) > 0$) as it does below $< X >$ (i.e., $(X - < X >) < 0$), which leads to $< X - < X >>= 0$. By defining the standard deviation as in Eq. 16, we make all values of $(X - < X >)$ to be positive or zero. We can make sense of the standard deviation (Eq. 16) as follows:

$$(X - < X >)^2 \qquad \text{(square of the distance between } X \text{ and } < X >) \tag{20}$$

and then

$$< (X - < X >)^2 > \qquad \text{(average of the square of the distance between } X \text{ and } < X >) \tag{21}$$

and thus

$$\sigma = \sqrt{< (X - < X >)^2 >} \qquad \text{(average "distance" between } X \text{ and } < X >\text{; square root "takes away" the square)} \tag{22}$$

note that above equation says "distance" with the quotes since σ is not exactly the average distance (the true average-distance is zero (by Eq. 19).

There is a convenient way to calculate the standard deviation. Note that

$$\sigma = \sqrt{< (X - < X >)^2 >} \tag{23a}$$
$$= \sqrt{< X^2 - 2X < X > - < X >^2>} \tag{23b}$$
$$= \sqrt{< X^2 > -2 < X >< X > - < X >^2} \qquad \text{(by Eqs. 14c and 15c)} \tag{23c}$$
$$= \sqrt{< X^2 > - < X >^2} \qquad \text{(convenient way to calculate } \sigma) \tag{23d}$$

(4) Variance: The variance is defined as σ^2:

$$\sigma^2 = \left\langle (X - < X >)^2 \right\rangle \tag{24}$$

The variance is useful because not having the square root in the standard deviation (Eq. 16) makes calculations simpler sometimes. In light of Eq. 23d, we have

$$\sigma^2 =< X^2 > - < X >^2 \tag{25}$$

(5) Fractional error: The fractional error is defined as

$$\frac{\sigma}{< X >} = \frac{\sqrt{\left\langle (X - < X >)^2 \right\rangle}}{< X >} \tag{26}$$

The fractional error is useful for determining how large the standard deviation is compared to the mean (i.e., as a percentage of the mean). Although not important for our purpose, there is also a related quantity called the **Coefficient of Variation (CV)**:

$$\text{CV} = \frac{\sigma^2}{< X >} \tag{27}$$

B. Continuous probabilities

We can define all the quantities above but now for continuous probabilities. The previous section dealt with **discrete probabilities**. There, we could count the total number of outcomes, N, which was a positive integer (i.e., $N = 1$, 2, 3, ...). This is because the random variable (the outcome) took on discrete values, such as in the case of rolling a six-sided dice. **Continous probabilities** for a random variable describe experiments whose outcome takes on a **continumm of values**. Here are two such examples.

Example A: With limited information, you use a microscope to observe a single cell. Can you predict at what time t after you begin observing the cell, it will divide?
Answer: Here, the time t is a random variable because you cannot exactly predict it with the limited information given to you (in fact, in real experiments, we never have enough information to predict typical behaviours of cells such as its division time t). But we can determine the probability that a cells divides time t after you begin observing it. Note that t can be any real number. That is, it can be any real number in the range, $0 < t < \infty$. So t can be 3.1415, or 3.14159, or 3141592, or 4.1, or 5, or 5.01, or 5.001, or 5.0001, and so on. You get the point. There is an infinite number of possible values for t. Of course, some values of t are less likely than others. For instance, intuition tells you that for a fast-dividing bacterial cell, t being a million years is very unlikely (and thus probability for that should be nearly, if not exactly, zero).

Example B: A point-sized particle is confined between two walls, one at $x = 0$ and the other at $x = L$ (L is the distance between the walls). It bounces back and forth between the two walls without loss of speed and it only moves along the x-axis only. While the particle is moving, you close your eyes and then you open them. Can you predict its position x before you open your eyes?
Answer: Here, the particle's position x is a random variable because you cannot definitely predict it without looking. But unlike in the previous example with t, the x is confined within a finite range: $0 \leq x \leq L$ (t was not confined to a finite range because t could be arbitrarily large). But x can still take on a continuum of values and, in fact, there is still an infinite amount of values of x. For example, $x = L/2$, $x = L/2 + 0.1 * L$, $x = L/2 + 0.01L$, $x = L/2 + 0.011L$, $x = L/2 + 0.0112L$, and so on. You get the point.

We just saw two examples of random variables whose values lie in a continuum. We now want to describe the probability for each outcome and then extend the definitions of statistical quantities so that they are defined for both discrete and continuous random variables. First, let's take example B and ask, "what is the probability that you find the particle between $x = 0$ and $x = L/2$ when you open your eyes?". According to our intuition, the probability should be $1/2$ (and it indeed is, as our calculation will show). We can also intuitively see, without any calculations, that the probability of finding the particle between $x = L/2$ and $x = L$ is also $1/2$ because the particle is moving back-and-forth without preferring one half of the room over the other. Note that

$$\frac{1}{2} = \frac{L/2}{L} \tag{28}$$

Following our intuition, the probability of x being $0 < x < L/4$ should be $1/4$ as is the probability of it being $L/4 < x < 2L/4$. Like wise, the probability of finding the particle between $x = 3L/4$ and $x = L$ should be $1/4$. Note that

$$\frac{1}{4} = \frac{L/4}{L} \tag{29}$$

We can see a pattern here. The probability of finding the particle between x_0 and $x_0 + \Delta x$ ($\Delta x > 0$) is

$$P((x_0, x_0 + \Delta x)) = \frac{\Delta x}{L} \tag{30}$$

where $P((x, x+\Delta x))$ denotes the probability of the random variable (position) x being within the interval $(x_0, x_0+\Delta x)$ for *any* value of x_0 (assuming that x_0 is properly sized so that $x_0 + \Delta x \leq L$). According to Eq. 30, the probability of finding the particle *exactly* at position x_0 is

$$P((x_0, x_0 + 0) = \frac{0}{L} = 0 \tag{31}$$

no matter what the value of x_0 is. This makes intuitive sense. It is simply saying that there are so many (in fact, infinitely many) values that x can take on within the range $[0, L]$ that the probability of x being *exactly* equal to x_0 (to 130 decimal places, if x_0 has exactly 130 decimal places), is zero. In other words, according to the definition of probability (Eq. 1),

$$P(x = x_0) = \frac{1}{N} = \frac{1}{\infty} = 0 \tag{32}$$

Note that the definition of probability of an outcome (Eq. 1) applies to both discrete and continuous random variables.

A convenient way to express Eq. 30 is by defining a **Probability Density Function (PDF)** $\rho(x_0)$ (Greek letter "rho"):

$$\rho(x_0) = \frac{1}{L} \qquad \text{(Probability Density Function (PDF) for Example B)} \tag{33}$$

which then lets us write Eq. 30 as

$$P((x_0, x_0 + \Delta x)) = \rho(x_0)\Delta x \tag{34}$$

Motivated by this example, we can generalize the concept of PDF to any situation, not just to a particle confined between two walls. For any situation, we define **probability density function (PDF)** for a random variable y to be $\rho(y)$ so that

$$P((y_0, y_0 + dy)) = \rho(y_0)dy \qquad \text{(General definition of Probability Density Function (PDF))} \tag{35}$$

where $P((y_0, y_0+dy))$ is the probability of the random variable y having a value in the infinitesimal interval (y_0, y_0+dy) (y_0 is some particular value). Note that, in general, $\rho(y)$ does not have to be a constant function. Some values of y may be more probable than others. As we will see, certain functional forms o $\rho(y)$ are given special names, such as **Poisson distribution**, **normal distribution**, and **uniform distribution**. For a finite interval $(y_0, y_0 + \Delta y)$, in which Δy is not infinitesimal like dy, Eq. 35 tells us that

$$P((y_0, y_0 + \Delta y)) = \int_{y_0}^{y_0+\Delta y} \rho(y)dy \qquad \text{(How to use a PDF)} \tag{36}$$

where $P((y_0, y_0 + \Delta y))$ is the probability of the random variable y being a value within (a potentially large) interval $(y_0, y_0 + \Delta y)$.

Normalization condition: The normalization condition (Eq. 10) for continuous PDF is

$$1 = \int_{y_{\min}}^{y_{\max}} \rho(y)dy \tag{37}$$

where $y_{\min}$ and $y_{\max}$ are minimum and maximum possible values of the random variable y. Note that they can be $\pm\infty$.

Definitions of statistical quantities:

For the most part, the definitions for statistical quantities that we gave for the discrete random variable are exactly the same for continuous random variables. But we repeat them here for completeness.

(1) Mean: The expectation value of a random variable y, given a continuous PDF $\rho(y)$, is

$$< y >= \int_{y_{\min}}^{y_{\max}} y\rho(y)dy \tag{38}$$

where $y_{\min}$ and $y_{\max}$ are minimum and maximum possible values of the random variable y.
(2) Standard deviation σ: Same as in the case of discrete random variables:

$$\sigma = \sqrt{< y^2 > - < y >^2} \tag{39}$$

(3) Variance σ^2: Same as in the case of discrete random variables:

$$\sigma^2 =< y^2 > - < y >^2 \tag{40}$$

The rest (fractional error and CV) are also exactly the same for both discrete and continuous random variables. We see above that the only difference is in how we calculate the mean for the two kinds of random variables (actually, when you learn more math (e.g., Dirac delta function), you will see that even the mean is computed in the same way for both discrete and continuous random variables - i.e. Eq. 38 and Eq. 13 are identical).

II. STOCHASTIC MODELS OF LIVING SYSTEMS - EXAMPLES

It is now time to use the mathematics of probabilities that we discussed above to biological systems. What we are about to do falls in the intersection of **statistical physics** (i.e., using probabilities to describe atoms and molecules) and **quantitative biology** (i.e., using math to describe biological systems). We will apply this to three settings: **(1) Berg-Purcell limit**, which is the limit to how accurately a cell or any receptor can determine the concentration of an external molecule, **(2) Luria-Delbruck analysis**, and **(3) Super-resolution microscope that defeats the diffraction limit (allows you to "see" molecules at the nanometer-scale using fluorescence**.

A. Berg-Purcell limit: Physics limits how accurately cells can sense their environment

In this section, we summarize a classic and one of biophysicists' favourite papers - "Physics of chemoreception" by Howard Berg and Edward Purcell in Biophysical Journal (1977). Here, Berg and Purcell asked and addressed whether there is a fundamental physical limit to how accurately a cell can detect the concentration of a diffusing molecule. They discovered, from first-principles calculations, that there is such a limit and this limit applies not just to living cells but to all non-living detectors, including any detection device that one may conceive of in the future. Experimentally, Howard Berg and others have verified this limit in the bacterium *E. coli* that senses concentration of desired molecules and swims towards it. This lower bound is also useful in understanding how accurately cells inside a developing embryo can sense a concentration of a **morphogen** - a chemical that cues an embryonic cell what type of a specialized cell it should differentiate into - and thus in understanding how accurately an embryo develops into a fully-formed organism (e.g., fly embryo developing into a fly). In this section, we will derive this limit.

Consider a cubic detector with a side length L (Fig. 1). This detector could be the entire cell, or a receptor inside a cell, or a location on DNA where a transcription factor should bind (then L is the length of that portion of DNA), or a receptor on the cell surface, just to list some of the many possibilities. The detector sits inside a much larger "bath" of volume V (i.e., $V \gg L^3$) (Fig. 1). Suppose that in this large bath, N molecules are diffusing around with a diffusion constant D. Let's assume that the N molecules are uniformly distributed inside the large bath. The average concentration $< c >$ is then

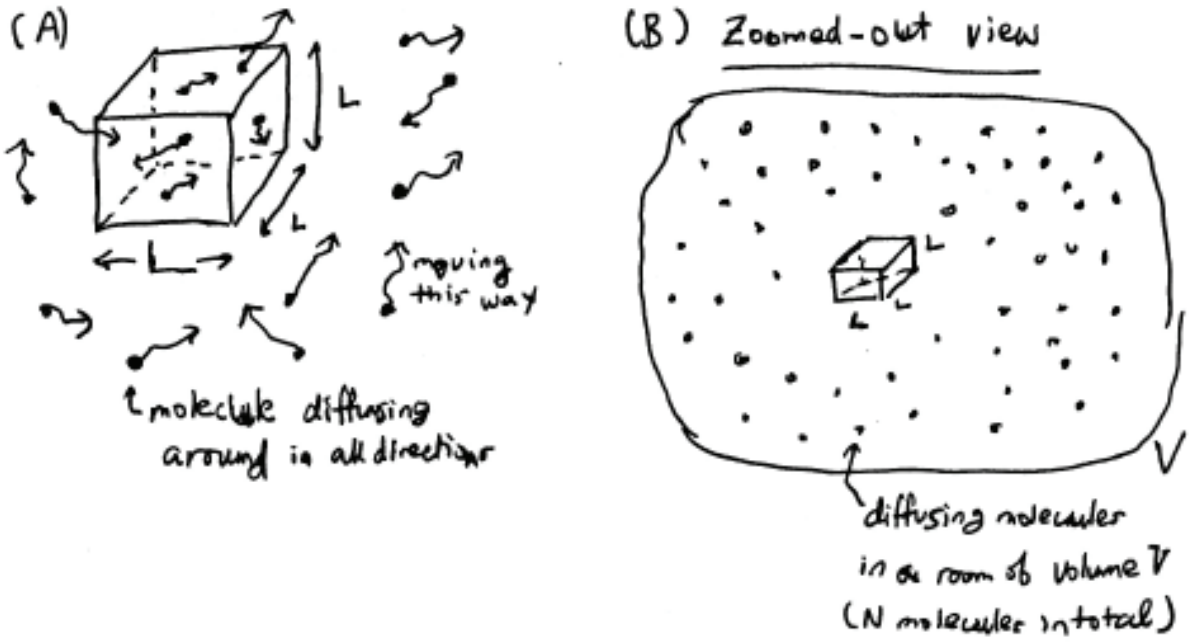

Figure 1. Setup for deriving the Berg-Purcell limit: N identical molecules are diffusing around in a large room of volume V. Inside the room is a cubic detector whose sides have length L and whose walls are completely permeable to the molecules. The detector "measures" how many molecules are inside it at a given time.

$$< c >= \frac{N}{V} \tag{41}$$

The main issue is that while we know the exact average concentration because we already *see* the N molecules inside the box in one snapshot, the cell does *not* know the total number of molecules inside the box. In fact, the cell may not encounter every one of the N molecules in a given amount of time because the molecules are diffusing (randomly moving) inside the box and thus some molecules may not *hit* the cell in a given time interval. The cell uses its detector to measure ("count") the number of molecules that come inside the detector. In this way, the cell samples a sub region in the large bath with its cubic detector, count the number of molecules that are inside the detector, and then from this deduces the concentration. On average, the number of molecules $< n >$ inside the cubic detector is

$$< n >=< c > L^3 \tag{42}$$

But this is the average value. If the cell measures the number of molecules x times, it will not obtain exactly the same value x times. Let's calculate how much variability there will be in the number n (random variable) among the different measurements. Let's consider one particular molecule. If we take a snapshot of the system, the probability p that we will find this particular molecule inside the cubic volume is

$$p = \frac{L^3}{V} \tag{43}$$

This is similar in logic to Example B above (the particle trapped between two walls). Now, suppose we paint a number on every one of the N molecules. So one molecule will have number "1" painted on it. The next one will have "2" on it. And so on. Now, out of these numbered molecules, let's consider a particular set of molecules - {1, 2, ..., n} (where $1 \leq n \leq N$). We now ask what the probability of finding this set of molecules within the detector is. Since each molecule is diffusing independently of each other, the probability of one molecule to be in the box is independent of what any of the other molecules are doing. Thus,

$$\underbrace{p \cdot p \cdot ... \cdot p}_{n \text{ times}} = p^n \qquad \text{(using Eq. 6)} \tag{44}$$

is the probability of finding the particular set of numbered molecules, {1, 2, ..., n}, in the detector. But this calculation also includes the possibility that the other molecules, to which we assigned a number larger than n, may also be present in the detector. We actually want a probability that only the molecules in the set {1, 2, ..., n} is in the detector and, at the same time, the other $N - n$ molecules being outside the detector. In our snapshot, the probability that the particular molecule is not inside the detector is 1-p. So, applying Eq. 6 and combining with Eq. 45, we have that

$$p^n(1-p)^{N-n} \tag{45}$$

is the probability of having only the set of numbered molecules, {1, 2, ..., n}, being in the detector and no one else. Now, there is nothing special about the set {1, 2, ..., n}. If we pick a different set of n numbered molecules, say {2, 3, 4, ..., n, n+1} or {N, N-1, N-2,..., N-n+1}, the probability of having exactly that particular set of molecules in the detector and no one else would still be Eq. 45. So, the probability $P(n)$ that exactly n molecules are inside the detector (which can be *any set* of n molecules that we pick from the N molecules) is

$$P(n) = \underbrace{p^n(1-p)^{N-n} + p^n(1-p)^{N-n} + ... + p^n(1-p)^{N-n}}_{\text{total number of groups of } n \text{ molecules}} \tag{46a}$$

$$= g \cdot p^n(1-p)^{N-n} \tag{46b}$$

where g is the total number of ways to generate groups of n molecules from a pool of N molecules. As we will see,

$$g = \binom{N}{n} \tag{47a}$$

$$= \frac{N!}{(N-n)!n!} \tag{47b}$$

which is called the **binomial coefficient**. Here, $k!$ (read "k factorial") means

$$k! = k(k-1)(k-2)...1 \qquad \text{(where } k \geq 1\text{)} \tag{48a}$$

$$\text{and} \quad 0! = 1 \qquad \text{(0! is specially defined like this)} \tag{48b}$$

If you have not encountered the binomial coefficient before, note that $N!$ is the total number of ways to arrange N molecules in a line. That is, if you assume that you paint a number on each molecule, from 1 to N, then you can have line up the molecules in the order [1, 2, 3,..., N], or in the order [2, 1, 3, 4, ..., N], or in the order [3, 1, 2, 4, ..., N], and so on. To count how many ways of forming the line there is, note that in the first location, we have N choices. Then after we picked the first molecule, we have $N - 1$ choices for the second position. So, there are $N(N-1)$ ways to form a line of length equal to two molecules (and thus a total of $N(N-1)$ *distinct* lines formed by two molecules). To count how many distinct lines of length equal to three there are, the same argument yields $N(N-1)(N-2)$ as the total number of such lines. Finally, there must be $N!$ distinct lines that N molecules can form. But for our detector, we are not asking about lines, in which we care about which molecule stands in front of which other molecule. We simply want to know, how many *sets* of n-molecules can be inside the cube. This means that we count an ordered line of [1, 2, 3,...., n] to be the same as an ordered line of [n, n-1. ..., 2, 1] since the same numbered molecules (1 to n) are in the box in both cases. This is where the binomial coefficient comes in. It eliminates the kind of double counting (overcounting) that we would have if we just counted the number of distinct ordered lines. Let's unpack Eq. **??** to understand it. First, note that

$$\frac{N!}{(N-n)!} = \frac{N(N-1)(N-2)...(N-n+1)(N-n)!}{(N-n)!} \tag{49a}$$

$$= N(N-1)(N-2)...(N-n+1) \tag{49b}$$

is the total number of ways to form a line of length equal to n with N molecules (see above paragraph if you don't understand this). But as we said before, we should treat some of these ordered lines as being the same (i.e., the cubic detector cares about sets of molecules instead of lines of molecules. Take one ordered line with n molecules. For example, let's look at [1, 2, 3, ..., n]. There are $n!$ ways of permuting these elements. Each permutation reorders the molecules' placement in the line but we have the same set of n molecules for each $n!$ permutation. Take another line with n molecules. For example, let's look at [2, 3, ..., n, n+1]. There are $n!$ ways of permuting these elements as well, each one giving a new line. But we also have the same set of n molecules for each of these lines. From this argument, we see that we have overcounted each set of n-molecules by $n!$ times. Thus, we must divide Eq. 49b (the total number of ordered lines of length equal to n) by $n!$, which yields the binomial coefficient:

$$\frac{N!}{(N-n)!n!} = \binom{N}{n} \tag{50}$$

Thus Eq. 46b becomes

$$P(n) = \binom{N}{n} p^n (1-p)^{N-n} \tag{51}$$

where $0 \leq n \leq N$. Eq. 51 is called the **binomial distribution** and we say that the random variable n is **binomially distributed**. The number of molecules in the box, n, is the random variable. We can calculate the average number of molecules, $< n >$, found in the detector at any snap shot in time from either the definition of the mean (Eq. 13) or, more simply, by noting that a binomial distribution describes an experiment with only two outcomes: success or failure (e.g., 'heads' or 'tail' like in a coin toss). The success probability is p and the failure probability is $1-p$. This is like throwing a (biased) coin with a probability p of getting a 'head' and a probability of $1-p$ of getting a 'tail'. If we throw such a coin N times, then we expect Np to be the number of times that we get a 'head' (indeed, this makes sense if $p = 0.5$). Thus, we can say that

$$< n >= Np \qquad \text{(for binomial distribution)} \tag{52}$$

The variance in n, denoted by σ_n^2, is

$$\sigma_n^2 = N(1-p)p \tag{53}$$

which we don't derive in this course (i.e., you can just accept it as it is for this course). We can rewrite Eq. 53 as

$$\sigma_n^2 =< c > V(1-p)p \qquad \text{(from Eq. 41)} \tag{54a}$$

$$=< c > V\left(1 - \frac{L^3}{V}\right)\frac{L^3}{V} \tag{54b}$$

$$=< c > L^3\left(1 - \frac{L^3}{V}\right) \tag{54c}$$

We actually want the variance associated with the measured concentration, σ_c^2, instead of the variance in the measured number, σ_n^2. We can convert σ_n^2 to σ_c^2 as follows:

$$\sigma_n^2 =< n^2 > - < n >^2 \tag{55a}$$

$$=< c^2 L^6 > - < c >^2 L^6 \tag{55b}$$

$$= L^6(< c^2 > - < c >^2) \tag{55c}$$

$$= L^6 \sigma_c^2 \tag{55d}$$

Substituting Eq. 55d into Eq. 54c yields

$$\sigma_c^2 =< c > \left(\frac{1}{L^3} - \frac{1}{V}\right) \tag{56a}$$

$$=< c > \left(\frac{V - L^3}{VL^3}\right) \tag{56b}$$

Assuming that $V \gg L^3$ (i.e., the bath has much larger volume compared to the detector), Eq. 56b becomes

$$\sigma_c^2 =< c > \left(\frac{1-(L/V)^3}{L^3} \right) \tag{57a}$$

$$\approx \frac{< c >}{L^3} \tag{57b}$$

This simple formula (Eq. 57b) is the variance in the measured concentration c. In other words, if the detector makes many independent measurements of the concentration inside it and then makes a histogram of all those measured concentrations, Eq. 57b is the variance in that histogram of c. The quantity that we are really interested in is not the variance itself but the **fractional error** $\sigma_c/ < c >$ associated with the measured concentration. We are interested in the fraction error because it tells us how "big" σ_c is (compared to the mean (actual) concentration $< c >$. "Big" or "small" are only meaningful in science when compared to some other quantity. The fractional error in c is

$$\frac{\sigma_c}{< c >} = \frac{1}{\sqrt{< c > L^3}} \tag{58}$$

This is the fractional error in the detector's measurement if the detector makes just one measurement of the concentration. But suppose that the detector makes M independent measurements of the concentration. It can then calculate the average of its M independent measurements, and from this deduce the concentration of the molecule. Intuitively, this tells you that the detector can more accurately determine the concentration. We can make this statement to be more quantitatively precise. The standard deviation σ_M associated with the average of M independent measurements is

$$\sigma_M = \frac{\sigma_1}{\sqrt{M}} \tag{59}$$

where σ_1 is the standard deviation when just the detector makes just one measurement (i.e., $M = 1$). From Eqs. 58 and 59, we see that the error in the detector's average of M independent measurements is

$$\frac{\sigma_c}{< c >} = \frac{1}{\sqrt{< c > L^3 M}} \tag{60}$$

How do we know how many measurements the cell will make before it averages those measurements? We cannot read the cell's "mind". But we can infer it from a different quantity that we can measure in experiments. Namely, suppose that we know that the cell has to determine the concentration within time interval T. Now, the question is how large can M be so that the cell can make M independent measurements within the time interval T. Clearly, the cell wants to make M to be as large as possible within the allotted time, according to Eq. 60. The key here is that the measurements must be *independent* of each other. This means that after the detector makes one measurement, it then must wait for all the molecules inside it to escape it, then wait for new molecules to enter, and then count those molecules inside. This way, the detector does not measure the same molecules in the next measurement. To see why the detector must wait until its inside is refreshed, note that if it makes the next measurement immediately after the current measurement, then it will measure the exact same value of concentration since none of the molecules inside it had time to move. In this case, the previous and the next measurements are *not* independent of each other. So we need to calculate the time it takes for the molecules inside the detector to escape. To estimate this, we use the fact that the diffusion constant D for the molecule has dimension of length^2/time. So we can roughly say that a molecule requires time $\tau = L^2/D$ to diffuse out of the detector (note that τ has a unit of time, so this makes sense dimension-wise). Then T/τ is the **maximum number of independent measurements** that the cell can make. In the next section, we will see how one can estimate T in an experiment. We can now rewrite Eq. 60, in experimentally accessible parameters, as

$$\frac{\sigma_c}{< c >} = \frac{1}{\sqrt{< c > LTD}} \qquad \text{Berg-Purcell limit} \tag{61}$$

Eq. 61 is the famous **Berg-Purcell limit**. The meaning of Eq. 61 is that if the cell has time interval T to deduce the concentration of some molecule inside it (if the detector is inside the cell) or outside it (if the detector is on the cell

membrane or if the entire cell itself is the detector), then the cell cannot determine the concentration with an accuracy higher than the fraction error stated in Eq. 61. Hence the Berg-Purcell limit is the **lower bound on accuracy** placed on the detector. It is remarkable that just by using sheer logic alone, we could deduce the fundamental limit to how accurately a cell or any sensor can measure a concentration of molecules.

One deficiency in our calculation of the Berg-Purcell limit is that we estimated but did not exactly calculate the M. Recent studies have proposed a method of exactly calculating the M. In doing so, these studies also suggested more elaborate mechanisms that ensure that each measurement is really independent. They include:

- K. Kaizu, W. de Ronde, J. Paijmans, K. Takahashi, F. Tostevin, and P.R. ten Wolde
  The Berg-Purcell limit revisited, *Biophysical Journal* (2014).
- W. Bialek and S. Setayeshgar
  Physical limits to biochemical sensing, *Proc. Natl. Acad. Sci. USA* (2005).

Despite the more elaborate calculations contained in these later studies, the bottom line is that the Berg-Purcell limit (Eq. 61) is still basically correct.

B. Applying the Berg-Purcell limit to the fruit fly embryo

In this section, we briefly summarize beautiful, physics-type high-precision experiments that Thomas Gregor and colleagues report in:

- T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek. Probing the limits to positional information, *Cell* (2007)

In this paper, Gregor et al. investigated how accurately nuclei inside the fruit fly embryo can measure the concentration of the morphogen called *Bicoid* that surrounds them. In the early embryo of the fruit fly, individual nuclei, not cells, are arranged next to each other in a near triangular lattice. The nuclei are not yet engulfed inside cell membranes - they are basically "cells" without membranes and other organelles. Simply put, the mother deposits the Bicoid mRNA at one end of the embryo. This end will form the head (i.e., *anterior*). Ribosomes that diffuse inside the embryo encounter and translate the mRNA into the Bicoid proteins. The Bicoid proteins diffuse from the head towards the other end of the embryo, which will form the tail (i.e., *posterior*). In so doing, a Bicoid-concentration gradient forms. Each nucleus in the embryo must "measure" the concentration of the Bicoid around itself. The concentration of the Bicoid that a nucleus measures is a function of the position of the nucleus relative to the embryo's head, where the Bicoid concentration would be the highest (since that is where the translation of Bicoid occurs - the mRNA is essentially localized at the embryo's head). Thus a nucleus can potentially "know" how far from the head it is by accurately measuring the Bicoid concentration. But due to the Berg-Purcell limit, there is a limit to how accurately each nucleus can measure the Bicoid concentration immediately surrounding it. Gregor et al. quantified how accurate is ?accurate enough? with calculations and tour de force, physics-type precision experiments on the fly embryo. For their experiments, Gregor et al. genetically engineered the fruit fly embryo in which each Bicoid protein is fused to a Green Fluorescent Protein (GFP) and built a high-resolution microscope called the **two-photon microscope** to measure the *number* of GFP molecules with a near integer-number precision (i.e., 1 GFP molecules, 2 GFP molecules, etc.). The latter was made possible by the fact that the two-photon microscope could resolve multiple Bicoid-GFP molecules that were very near each other as distinct molecules.

We now discuss the paper's main findings. Two hours after the fruit fly egg has been fertilized, the length L of the embryo is approximately 500 μm and the average distance Δx between the centers of two adjacent nuclei (which we treat as spheres) that are aligned along the straight line joining the head to tail (anterior-posterior) is approximately 8 μm (see Figure in the paper). Each nucleus has a diameter of approximately 5 μm. Based on these measurements, we have

$$\frac{\Delta x}{L} = \frac{8\ \mu\text{m}}{500\ \mu\text{m}} \approx 0.016 \tag{62}$$

Thus having two adjacent nuclei being able to tell that they are at two different locations along the anterior-posterior axis means that the nuclei must be able to resolve a difference in length by about 1.6% of the embryo's total

length. Suppose that a nucleus is at position x where the Bicoid concentration is $c(x)$. Let $\Delta c + c(x)$ be the Bicoid concentration on its adjacent nucleus. Suppose that the Bicoid protein, after it is made at the anterior end of the embryo (i.e., at $x = 0$), diffuses freely inside the embryo towards the posterior end. Then solving the 1-dimensional diffusion equation with a source of molecules at the anterior yields

$$c(x) = c_0 e^{-x/\lambda} \tag{63}$$

where λ is the diffusion length. Gregor et. al. have measured the concentration of the Bicoid inside tens of fly embryos. In every embryo, they found that the concentration $c(x)$ indeed exponentially decayed as in Eq. 63 from the anterior to posterior. Intriguingly, they found that there was very little variability in the concentration profile $c(x)$ from one embryo to another, despite coming from different mothers (in a later study, Gregor's lab has shown that all mother flies likely lay nearly the same, integer number of Bicoid mRNA at the anterior ends of their embryos!). Their measurements showed that the diffusion length λ is approximtaely 100 μm. For two adjacent nuclei, we have

$$\left|\frac{\Delta c}{c(x)}\right| = \left|\frac{1}{c(x)}\frac{dc}{dx}\right|\Delta x \tag{64a}$$

$$= \frac{\Delta x}{\lambda} \tag{64b}$$

$$= \frac{8\ \mu\text{m}}{100\ \mu\text{m}} \tag{64c}$$

$$\approx 0.10 \tag{64d}$$

According to Eq. 64d, the Bicoid concentration on a given nucleus is different by approximately 10% from the Bicoid conentration on its adjacent nucleus. Thus for two adjacent nuclei to "know" that they are apart by a distance of Δx, they must be able to distinguish 10% or less of a difference in their respective Bicoid concentrations.

C. Luria-Delbruck analysis

Here's another, classic application of stochastic model to living systems. This time, we're interested in knowing how a cell inherits a trait from its mother cell. In a classic experiment, done in 1943, Salvador Luria (biologist) and Max Delbruck (physicist) used statistical modelling to determine how some E. coli cells develop resistance to a virus, called bacteriophage, that infects them. To a non-resistance E. coli cell, bacteriophage infection is lethal. This is because a bacteriophage injects its RNA into a non-resistant cell which then translates the RNA into proteins, which then combine to form more bacteriophages inside the bacteria. The bacteriophages then burst open the E. coli cell, killing it, are released into the environment. They then go and find more bacterial cells to infect. This is how bacteriophages and viruses, in general, replicate themselves - they use host cells. But some E. coli cells were resistant to bacteriophages. Bacteriophages could not inject their RNA into these cells. But a mystery was how non-resistant cells arose within a population. Researchers had observed that in a population of seemingly identical E. coli cells that are all living in the same environment, only a fraction of the cells became resistant. Prior to Luria and Delbruck's work, researchers proposed two possible mechanisms for how the resistance trait arose: (1) **Lamarckian inheritance** and (2) **Darwinian inheritance**. In the Lamarckian picture, the stress of encountering bacteriophages *induced* resistance in some of the cells in the population. In the Darwinian picture, resistant cells *pre-existed before* they encountered the bacteriophages because of spontaneously formed within a population. In a modern language, we would now say that these cells arose from spontaneous mutations in certain genes that conferred resistance to the cells. In 1943, when Luria and Delbruck published their experiment, very little was known about how daughter cells actually inherited traits from their mother cells. After all, this was nearly 10 years before Watson's and Crick's discovery of the structure of DNA which conclusively established that DNA was the molecule that carried the genetic information. While one might be tempted, now in hindsight, say that the Darwinian mechanism is the one that yielded the phage-resistant E. coli cells - and it is indeed - recent studies have discovered that cells (e.g., cancer cells) can develop resistance to a stress (e.g., cancer drug) through the Lamarckian mechanism. Luria and Delbruck determined that the phage-resistant cells arose through the Darwinian mechanism. Here we give a brief summary of how they determined this (schematic details are in the lecture slides).

The main logical puzzle that Luria and Delbruck faced was that in order to determine which cells were resistant to bacteriophages and how many of them were resistant within a population, they had to first subject the E. coli

cells to the bacteriophages. While this could then tell them how many cells were resistant (by counting the total number of cells that survived the phage-attack), they could not tell whether or not these cells were resistant *before* they encountered the phages. To do so, it seemed that one had to measure some trait in the bacterial cells *before* they encountered the phages. But not knowing what molecules or genes caused the resistance - and keep in mind that the concept of genes residing on DNA that gave rise to traits was not a popular idea back in 1943 - there was no way that Luria and Delbruck could any such measurement before giving phages to the bacteria. So instead, they had an ingenious idea. Both the Lamarckian and the Darwinian mechanisms are *stochastic* - both cause only a fraction of the cell population to become resistant. This means that if we had 1000 identical populations that all start with the same number of cells, then the number of survivors (resistors) after they all face a phage attack would vary from population to population. Luria and Delbruck's ingenuity was that they realized that by measuring the variability in the number of survivors from the replicate populations, they could distinguish the two mechanisms (see schematic details in the lecture slides).

To see how this would work, let's first consider the Lamarckian model. Suppose that p is the probability that a cell becomes resistant after we introduce bacteriophages to the cell population. If N is the total number of cells in the population, then the probability of having n resistors after introducing bacteriophages is

$$P_(n; p, N) = \binom{N}{n} p^n (1-p)^{N-n}, \tag{65}$$

which is the binomial distribution. This makes sense because, for each cell, we are throwing a (biased) "coin" whose probability of getting one side (resistant) is p and the other side is thus $1-p$ (non-resistant). As you will show in a problem set, in the limit of the population size N becoming very large (i.e., $N \to \infty$), the above binomial distribution becomes the Poisson distribution:

$$P(n; \mu) = \frac{\mu^n}{n!} e^{-\mu} \tag{66}$$

Now, suppose that we have 1000 identical populations that all start with the same number of cells and that we infect all of them with the same amount of bacteriophages. Suppose that we leave all 1000 populations in the presence of the bacteriophages for the same amount of time t. After time t, we kill all the bacteriophages with a drug. We then transfer each liquid culture containing the surviving cells (i.e., a population of cells is grown in a well-mixing liquid media) to an agar plate on which the cells can form colonies. Importantly, we spread out the cells on the agar pad so that one colony arises from a single surviving cell. Then the number of colonies that form on the agar pad is the number of resistant cells that existed in the liquid culture after time t. By counting the colonies on each plate - one plate per population - and using the fact that they knew the average doubling time of the cells, they could determine the number of survivors n that existed in each population *just after* phages started attacking them. If the Lamarckian mechanism was responsible for creating resistors, the n would follow a Poisson distribution (Eq. 66).

Analytically deriving the probability distribution for n that describes the Darwinian mechanism is more difficult but feasible. That is what Luria and Delbruck did. But it is easier to computationally simulate the distribution using, for instance, MATLAB. You will do this in a problem set. Luria and Delbruck then compared the experimental data - a histogram of the number of survivors from each population - with the expected distributions from each of the two models. In doing so, they discovered that the Darwinian distribution, for the same mean number of survivors as the Poisson distribution, has a much longer tail (i.e., larger number n of survivors) than the Poisson distribution. For this reason, the Darwinian distribution is called the **jackpot distribution** (like winning a lottery or a game at a casino) and correctly concluded that the Darwinian mechanism produced the phage-resistant E. coli cells.