1. Statistics and differential equations review

Recall that with n trials and per-trial probability of outcome p, the probability of observing $k = 0, 1, 2, \ldots$ outcomes is

$$\Pr(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{the binomial distribution}$$

And recall that if an event happens with rate λ per unit time, the probability of observing $k = 0, 1, 2, \ldots$ outcomes in unit time is

$$\Pr(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad \text{the Poisson distribution}$$

The mean of a discrete distribution is given by

$$\mu = \sum_{k \in \text{all outcomes}} k \Pr(k)$$

and the variance by

$$\sigma^2 = \sum_{k \in \text{all outcomes}} (k-\mu)^2 \Pr(k)$$

The standard deviation σ is the square root of the variance, and has the same unit as the mean.

For a binomial distribution, $0 \leq k \leq n$, whereas for a Poisson distribution, $0 \leq k \leq \infty$.

R1. You flip what you think is a fair coin ten times and get heads every time.
   (a) (1pt) What is the probability of this result if the coin's true probability of landing heads $p = 0.5$?
   (b) (1pt) What p would make this outcome (ten out of ten heads) have probability 0.01?
   (c) (1pt) What is the mean of a binomial distribution with n trials and outcome probability p? Simplify as much as possible.

R2. Your favorite bacterium generates a Poisson-distributed number of progeny with $\lambda = 1.5$ average number of progeny per hour.
   (a) (1pt) What is the probability that a bacterium generates zero offspring in an hour?
   (b) (1pt) What is the probability that it generates at least one offspring?
   (c) (1pt) What are the mean and standard deviation of the number of offspring per hour, both in terms of λ and in actual numbers for this bacterium?

R3. A yeast cell buds to produce a daughter cell every two hours. Assume that, in a population of 10^8 cells synchronized to bud simultaneously, this process is well-described by the function $n(t+1) = Rn(t)$ where $n(t)$ is the number of cells at time t (in hours) and R is the rate of growth.
   (a) (1pt) What is R?
   (b) (1pt) If we now ask for the instantaneous rate of growth r rather than the per-unit-time rate of growth, we have that $\frac{dn(t)}{dt} = rn(t)$. What is r in terms of R? Hint: write out the solution for $n(t)$ in each case.

2. Evolution of antibiotic resistance

1. As discussed in class, organisms struggle for survival against statistics as well as each other.

   Download the `expgrowth.R` script from the Canvas site (`Assignments/HW1/src`) and run it in RStudio. Manipulate the population parameters near the top: `n0`, the number of starting organisms, and `max.gen`, the number of generations to track.

   (Feel free to explore further and alter the fitness and so on.)

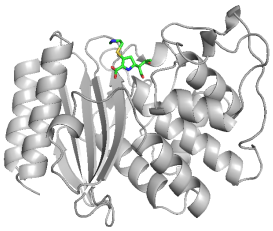
   (a) (2pts) Run 100 replicates for 100 generations with initial population sizes of $n_0 = 1$, 10, 100, and 1000. At each n_0, after 100 generations, what fraction of the populations have died out (*i.e.*, have $n = 0$)?

   (b) (2pts) Now run 100 replicates with initial population sizes of $n_0 = 1$, 10, 100, and 1000, but run for n_0 generations (1 generation for $n_0 = 1$, 100 generations for $n_0 = 100$, *etc.*). What do you notice?

2. In β lactamases, several sites commonly accumulate changes that alter resistance to various β-lactam antibiotics. We would like to know where these residues are relative to the antibiotic itself. Download the `TEM1-imipenem-from-1BT5.pse` file from the Canvas site (`Assignments/HW1/data`). Do a little preparation: show the TEM-1 molecule as a gray cartoon on a white background, show the β-lactam inhibitor imipenem, and orient the view:

```
load your_file_path/TEM1-imipenem-from-1BT5.pse
hide everything
bg_color white
show cartoon, TEM-1
color gray70, TEM-1
orient TEM-1
show sticks, imipenem
```

   Using your mouse, rotate the protein until you've got a good view of the inhibitor; something like this:



   (a) (2pts) Select residues M69, S70, R164, M182 and G238, which have been inferred to play important roles in the evolution of drug resistance:
   `select resi 69+70+164+182+238`
   Show these residues as red spheres:
   `show spheres, sele; color red, sele`
   Raytrace the result and print/paste it into your answers:
   `ray; png your_file_path/tem1-residues-highlighted.png`

(b) (2pts) Which of these residues is closest to the substrate? Which is farthest away?

3. Highly conserved amino acid sites are often assumed to be functionally important. Let's use Python to study the diversity at amino acid sites in TEM-1 β-lactamases.

   Download the `CalculateNumAaPerSite.py` script and the `tem1-alignment.fasta` file from the Canvas site. If you haven't done so already, read the "Programming and Installation" guide document we've written, install Python and the required modules (biopython), and organize your homework directory structure accordingly.

   The multiple sequence alignment contains 125 TEM-1 β-lactamase sequences, each of which is 287 amino acids long.

   Open up `CalculateNumAaPerSite.py` in Spyder. Change the `base_directory` variable to your homework directory's path. Note that you need to have your directories set up correctly, as we detail in the Programming Tips guide. Read the script and comments carefully to understand what it will do. Run the script.

   (a) (1pt) Examine the output histogram `SiteDiversityHist.png` in your `hw1/results` folder. How many invariant sites did you find?

   (b) (2pts) Which sites are invariant? Modify the script to identify these invariant sites.

   (c) (1pt) Return to PyMol, as in the previous question. Show these invariant sites as blue spheres. Print/paste the result.

   (d) (2pts) What is the relationship between the absolutely conserved sites and those found to be functionally important in the evolution of resistance?

4. Hydrolysis of β-lactam antibiotics allows bacteria to grow. Suppose we have the chemical reaction describing the spontaneous hydrolysis of ampicillin: $\mathrm{A} \xrightarrow{k} \emptyset$ with $k = 0.01\ \mathrm{s}^{-1}$.

   (a) (1pt) On average, how fast does 10 μM ampicillin degrade spontaneously (in μM per second) at the beginning of the reaction?

   (b) (1pt) Plot $[A](t)$ as a function of time in R (see tip below).

   (c) (1pt) Suppose a β-lactamase L catalyzes this hydrolysis: when bound to β-lactamase, ampicillin hydrolyzes 10,000 times faster than its spontaneous rate. What is k_{cat}?

   (d) (2pts) Suppose we have the full scheme $\mathrm{A} + \mathrm{L} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{A} \cdot \mathrm{L} \xrightarrow{k_{\mathrm{cat}}} \mathrm{L} + \emptyset$, with $k_{\mathrm{on}} = 1000\ (\mu\mathrm{M} \cdot \mathrm{s})^{-1}$, $k_{\mathrm{off}} = 10\ \mathrm{s}^{-1}$. What is the value of the Michaelis constant K_M for this reaction?

   (e) (2pts) On average, how fast is the degradation of A at the beginning of a reaction that again starts with an initial concentration of 10 μM ampicillin, along with an effective concentration of 1 μM β-lactamase?

TIP: In R, making a plot is easy:

```
> t <- 1:100
> k <- 0.1
> plot(t, exp(k*t), type="l", col="red",
+      xlab="Time (s)", ylab="Amount")
```