BIO 5329: PROBLEM SET 2

NATHAN BAKER

(BAKER@CCB.WUSTL.EDU)

Due at the start of class on October 4. Please hand in the analytical problems in class and e-mail the *well-commented* Matlab M-files to me (baker@ccb.wustl.edu) before class.

1. Binding curve analysis (20 points)

Ligand binding to receptors is often represented in terms of series called “binding polynomials.” This question deals with the manipulation of such series.

a. (5 points) Suppose we have a single species X of ligand that can bind to N (non-interacting) sites of a biomolecule with a binding polynomial of the form:

$$Q_X = \sum_{n=0}^{N} K^n x^n \tag{1.1}$$

where K is the stoichiometric association constant for each site and x is the activity of the ligand (i.e., effective concentration). Note that these definitions imply $K, x \geq 0$. Show that

$$Q_X = \frac{(Kx)^{N+1} - 1}{(Kx) - 1}. \tag{1.2}$$

b. (5 points) The degree of binding (moles ligand bound/moles biomolecule) can be determined from Q_X by the formula

$$\overline{X} = \frac{\partial \ln Q_X}{\partial \ln x}. \tag{1.3}$$

Determine the degree of binding from the Q_X expression you derived above.

c. (5 points) Suppose $N \to \infty$ and $Kx < 1$. Find analytic expressions for Q_X and $\overline{X}$ under these assumptions.

d. (5 points) Let

$$Q_{XY} = \sum_{m=0}^{N-n} \sum_{n=0}^{N} K^{(m+n)} y^n x^m \tag{1.4}$$

be the binding polynomial for a system of "identical linkage"; e.g., where molecules of species X and Y bind with the same affinity to N identical sites; however, X and Y cannot bind to the same site simultaneously. Find a closed-form expression for Q_{XY} (e.g., with no $\sum$s) analogous to Eq. 1.2 for Q_X.

2. Series manipulation (20 points)

Find the closed form for the following series; be sure to *show your work*. Assume all series are being evaluated for arguments where they are convergent (i.e., you can perform arithmetic and calculus on the series).

a. (5 points) Evaluate

$$f(x) = \sum_{n=0} \frac{(\imath x)^n}{n!} + \sum_{n=0} \frac{(-\imath x)^n}{n!} \tag{2.1}$$

where $\imath = \sqrt{-1}$.

b. (5 points) Evaluate

$$f(x)=\sum_{n=0}^{\infty}\frac{x^{2n}}{n!}. \tag{2.2}$$

c. (5 points) Use the answer you just calculated above to evaluate

$$g(x)=\sum_{n=0}^{\infty}\frac{(-1)^n x^{2n+1}}{n!(1+2n)}. \tag{2.3}$$

d. (5 points) Evaluate

$$f(x)=\sum_{n=0}^{\infty}\frac{(-1)^n x^{2n+1}}{2n+1} \tag{2.4}$$

3. A LIMIT OF LARGE NUMBERS (20 POINTS)

Suppose we have an ion channel which is stochastically gated; for independent observations, the probability of finding the channel open is p and the probability of finding it closed is $1-p$. If we make N observations, the probability of finding the channel open n times is

$$P_N(n)=\frac{N!}{n!(N-n)!}p^n(1-p)^{N-n}. \tag{3.1}$$

a. (5 points) Use Stirling's approximation

$$\ln N! \sim N\ln N - N. \tag{3.2}$$

to write an expression for $\ln P_N(n)$ at large N and n.

b. (5 points) Find the value n^* which maximizes your large-N approximation of $\ln P_N(n)$.

c. (5 points) Expand your large-N approximation for $\ln P_N(n)$ to second order (i.e., include only terms up to n^2) around the maximum you identified in the previous step.

d. (5 points) Normalize $P_N(n)$ given the approximations above such that

$$\int_{-\infty}^{\infty} \mathcal{N} P_N(n) dn = 1 \tag{3.3}$$

where $\mathcal{N}$ is a multiplicative constant (the normalization constant) you need to define. What is the probability of observing $n > M$ (where M is a big positive number) opening events?

4. Binomial distribution confidence intervals (20 points)

Let's revisit another version of the ion channel problem above. Suppose you're able to make a series of N measurements to determine if the channel is open or closed. For those N measurements, you observe the channel in the open state N_o times and thus estimate the probability of the channel being open as $\hat{p}_o = N_o/N$.

Unfortunately, your grant budget was recently cut by the NIH and you can only afford to perform a small number of experiments such that N is small. In addition to the stress associated with NIH funding, you now have to figure out how to analyze the error in your $\hat{p}_o$ estimate for this small number of measurements. Fortunately, there are two formulæ to help in this task. The upper α (e.g., for $\alpha = 0.95$, 0.99, etc.) confidence interval on $\hat{p}_o$ is p_u and given by the equation

$$\sum_{k=0}^{N_o} \binom{N}{k} p_u^k \left(1 - p_u\right)^{N-k} = \frac{\alpha}{2} \tag{4.1}$$

where the binomial coefficient is defined as

$$\binom{a}{b} = \frac{a!}{(a-b)!b!}. \tag{4.2}$$

Likewise the lower confidence interval p_l is governed by the equation

$$\sum_{k=0}^{N_o-1} \binom{N}{k} p_l^k \left(1-p_l\right)^{N-k} = 1 - \frac{\alpha}{2}. \tag{4.3}$$

Your job is to write Matlab code to determine $\hat{p}_o$, p_u, and p_l for $\alpha = 0.99$ (99% confidence interval) for the following experimental results:

Exp	A	B	C	D
N_o	1	2	4	20
N	5	10	20	100

Hints:

- $0 \le p_l < \hat{p}_o < p_u \le 1$
- Use `fminbnd` (on an alternative version of the above equations...) rather than `fzero`.

5. INTEGRATION (20 POINTS)

Recall that, given a set of quadrature points $x_1, \ldots, X_N$ with even spacing $x_{i+1} - x_i = h$, we can represent the integral of a function $f(x)$ by

$$\int_{x_1}^{x_N} f(x)dx = \sum_{i=1}^{N} w_i f(x_i). \tag{5.1}$$

For the trapezoidal quadrature rule, the weights are given by

$$w_i = h \begin{cases} \frac{1}{2} & i = 1 \\ 1 & 1 < i < N \\ \frac{1}{2} & i = N. \end{cases} \tag{5.2}$$

For Simpson's quadrature rule, the weights are given by

$$w_i = \frac{h}{3}\begin{cases} 1 & i = 1 \\ 4 & 1 < i < N \text{ and } i \text{ even} \\ 2 & 1 < i < N \text{ and } i \text{ odd} \\ 1 & i = N. \end{cases} \tag{5.3}$$

Use Matlab to evaluate

$$I = \int_{-3}^{3} e^{2\sin(2\pi x)} dx \tag{5.4}$$

with $N = 5, 10, 20, 40, 80, 160$ quadrature points using both the trapezoidal and Simpson's rules. Compare your result with the true integral

$$I = \int_{-3}^{3} e^{2\sin(2\pi x)} dx = 6 I_0(2) \approx 13.67751181402, \tag{5.5}$$

where $I_n(x)$ is the n-th order modified Bessel function of the first kind. Which quadrature rule converges to the true value most quickly?