

*Stochastic Gene Expression Project: Homework 3*  
*Due February 1st at 2:15 PM*

**Problem 1: Kullback-Leibler Divergence.**

In order to decide whether additive or multiplicative noise better approximates the true Gillespie noise of our gene regulation systems, we need some way to compare probability distributions. One way is using the so-called *Kullback-Leibler divergence* (KLD). For two continuous (one-dimensional) probability distributions  $p$  and  $q$ , the KLD is defined as

$$D_{KL}(p||q) := \int_{-\infty}^{\infty} p(x) \log \left( \frac{p(x)}{q(x)} \right) dx .$$

Since computers can't hold entire continuous probability distributions in their memory, we instead work with our probability distributions evaluated at a discrete set of points  $x_i$ , so that we have a set of  $p_i := p(x_i)$  and  $q_i := q(x_i)$ , with  $\sum_i p_i = \sum_i q_i = 1$ . For discrete probability distributions, the KLD is defined as

$$D_{KL}(p||q) := \sum_i p_i \log \left( \frac{p_i}{q_i} \right) .$$

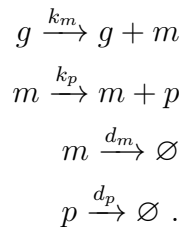
In this problem, you will show that the KLD is nonnegative. Assume that if  $p_i \neq 0$  for some  $i$ , then  $q_i \neq 0$ .

---

- (a) Show that  $\sum_i p_i \log \left( \frac{p_i}{q_i} \right) = - \sum_i p_i \log \left( \frac{q_i}{p_i} \right)$ .
- (b) Use the inequality  $-\log x \geq -(x - 1)$  and the result from part (a) to show that  $\sum_i p_i \log \left( \frac{p_i}{q_i} \right) \geq \sum_i p_i - q_i$ .
- (c) Recall that the sum only contains  $i$  for which  $p_i \neq 0$  (if  $p_i = 0$  for some  $i$ , then the term obviously vanishes). This means that it is possibly missing indices  $j$  for which  $q_j \neq 0$ . Argue that this means  $\sum_i p_i - q_i = 1 - \sum_i q_i \geq 0$ . Doing this proves that the KLD is nonnegative.
- (d) Suppose  $p_i = q_i$  for all  $i$ . Show that the KLD is zero.

## Problem 2: Additive and multiplicative noise via Taylor expansion.

Consider the unregulated gene, whose defining reactions are



In this problem, you will compute the optimal additive and multiplicative noise approximations to the Gillespie noise function.

---

- (a) Doing the usual QSS approximation for  $m$ , write out the SDE for  $p$ .
- (b) Show that the noise function is equal to  $\sqrt{d_p}\sqrt{p_{ss} + p}$ .
- (c) Assume that  $p$  is close to its steady state value  $p_{ss}$ , so that  $p = (1 + \epsilon)p_{ss}$  for some small  $\epsilon > 0$ . Show that the noise function is equal to  $\sqrt{2d_p p_{ss}}\sqrt{1 + (\epsilon/2)}$ .
- (d) Use the fact that  $\sqrt{1 + x} \approx 1 + (1/2)x$  for small  $x$  and the definition of  $\epsilon$  to show that the noise function is equal to  $A + Bp$  for some constants  $A$  and  $B$ . What are  $A$  and  $B$ ? Which one corresponds to the additive part of the noise, and which one corresponds to the multiplicative part?
- (e) To find the optimal additive noise approximation, suppose that  $p \approx p_{ss}$  in the noise function, so that it equals  $\sigma_a := A + Bp_{ss}$ . Simplify the expression. You should find  $\sigma_a = \sqrt{2d_p p_{ss}}$ .
- (f) To find the optimal multiplicative noise approximation, incorporate  $p$  into the  $A$  term by writing  $A \rightarrow (A/p_{ss})p$ . Calculate  $\sigma_m := (A/p_{ss}) + B$  and simplify the expression. You should find  $\sigma_m = \sqrt{(2d_p)/p_{ss}}$ .

### **Problem 3: Preliminary two gene model code.**

Look over the preliminary code in this week's GitHub folder to familiarize yourself with the current state of the two gene model. Let me know if you have any questions about its organization.