

Optimal inference of environmental conditions

1. Analytic results for the Monod model.

- Use `GillespieSwain.m` to simulate $N = 10$ stochastic trajectories of the Monod model (with the parameterisation above) for sugar copy number 0, 1, 10, 100.
- Obtain the average occupancy of the B state for each of these copy numbers.
- Compare the analytic result:

$$R = \frac{(1 + S/K_R)^n}{L(1 + S/K_T)^n + (1 + S/K_R)^n} \quad (13)$$

with your data.

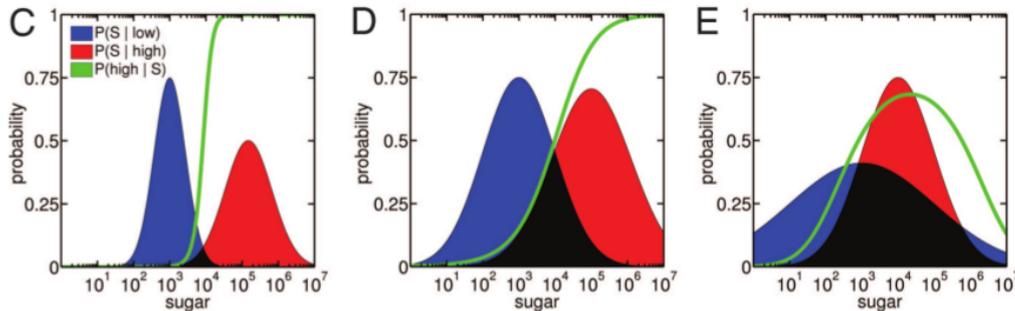


Figure 1: Different $P(S|\text{high})$, $P(S|\text{low})$ and resulting $P(\text{high}|S)$ illustrated in *Noisy information processing through transcriptional regulation*, Libby *et al.*, Proc. Natl. Acad. Sci. USA **104** 7151 (2007). Bonus points for the first to tell me what's dodgy with these plots.

2. Modelling high and low sugar states.

Now we'll consider the stochastic chemical situations corresponding to different biological environments. Say that there are two different environments that our cell may experience: high and low sugar states. The distribution of sugar in each state is given by $P(S|\text{high})$ and $P(S|\text{low})$; the probability $P(\text{high}|S)$ (which we're interested in in our inference problem) depends on these (see Fig. 1).

- If we have an evenly split prior belief that we may be in either state, and observe a given S measurement, use Bayes' rule to write down the probability $P(\text{high}|S)$ that we are in the high-sugar state.
- The file `LogNormal.m` produces log-normal pdfs that we can use to represent the probabilities $P(S|\text{high})$ and $P(S|\text{low})$. Use the parameters $\mu_{\text{low}} = 50$, $\sigma_{\text{low}} = 20$, $\mu_{\text{high}} = 100$, $\sigma_{\text{high}} = 50$, and plot $P(\text{high}|S)$ against S .

3. Optimally inferring sugar state with stochastic biochemistry.

Ideally, we would like our chemical system to yield a B state probability that exactly matches $P(\text{high}|S)$; thus accurately inferring the likelihood that we are in a high-sugar state. The file `FunctionToFit.m` implements the analytic result above.

- Use MATLAB's `lsqcurvefit` function to fit your previously computed $P(\text{high}|S)$ with an optimal parameterisation of the Monod model. You may need to increase the maximum allowed number of function evaluations `MaxFunEvals` in the optimisation process to do this.
- Plot the occupancy of the binding protein state (the response of our inference system) against $P(\text{high}|S)$ (the theoretical response from Bayesian inference).
- What factors should be taken into account when assessing the practical ability of biological systems to behave as this theoretical treatment would suggest? Can you think of ways that the system could be modified to provide a better and/or more robust fit to the theoretical Bayesian behaviour? Hint: remember the effects of n from the last practical – try implementing this.