T-61.5110 Modeling biological networks

Exam, December 20, 2013

You are NOT allowed to use calculators or any other additional equipments/material in the exam. Please write your answers in English. Please write carefully.

1. Consider the prokaryotic auto-regulation model in Figure 1, which is taken from the course book: g=gene, r=transcript, P=protein, $P2$=protein dimer complex (formed of two proteins P), $RNAP$=RNA polymerase, p=binding/operator site of $RNAP$, q=binding/operator site of P2; RNAP can transcribe the gene g unless $P2$ blocks the transcription. Construct the corresponding coupled chemical reactions (i.e., reaction network model). Also formulate the model as a Petri net $(P, T, Pre, Post, M)$ using the so-called matrix formalism. (6 points)

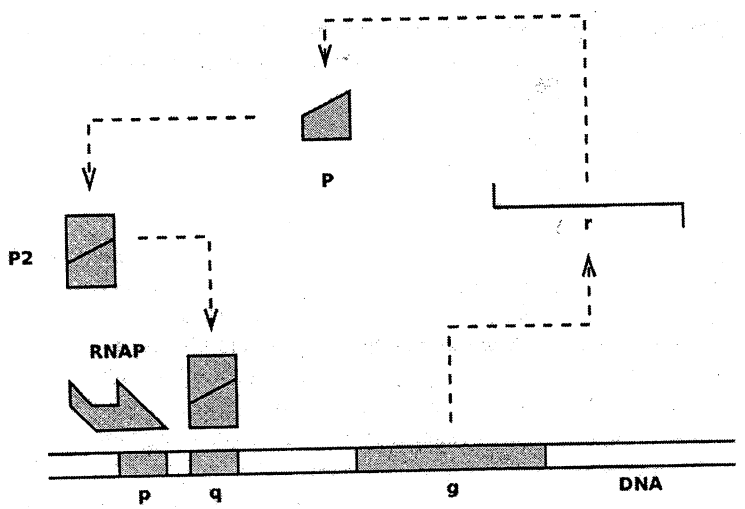

Figure 1: A simplified prokaryotic auto-regulation model.

2. Consider the simple model of Michaelis-Menten enzyme kinetics

$$\begin{aligned} S + E &\rightarrow SE \\ SE &\rightarrow S + E \\ SE &\rightarrow P + E \end{aligned}$$

   a) Represent it mathematically as a Petri net, $N = (P, T, Pre, Post, M)$, assuming that there are currently 100 molecules of substrate S, 20 molecules of the enzyme E, and no molecules of the substrate-enzyme complex SE or the product P. (2 points)

   b) If the first reaction occurs 20 times, the second 10, and the last 5, what will be the new state of the system? (2 points)

   c) Can you identify any P-invariants (i.e., conservation laws) for this system? If you can, identify one P-invariant. (2 points)

3. Assume you are given a PSFM model θ for a DNA binding protein A. Assume the width of the PSFM (and thereby the width of the binding site/motif) is n.

   a) You are given two DNA sequences, a wild-type $X = (x_1 x_2 x_3 \dots x_n)$ and a mutant $X^* = (x_1 x_2^* x_3 \dots x_n)$, where $x_2 \neq x_2^*$. Describe how you can computationally decide which one of the sequences, X or X^*, is more likely to bind protein A. (3 points)

   b) You are given a DNA sequence $Z = (z_1 z_2 \dots z_\ell)$, where $\ell > n$. Explain how you can assess the statistical significance that Z contains a binding site for protein A. (3 points)

4. a) Explain the Gillespie algorithm for simulating coupled chemical reactions. You can assume general stochastic rate constants $c_1, \dots, c_v$ and hazard functions $h_1(\mathbf{x}, c_1), \dots, h_v(\mathbf{x}, c_v)$ for all reactions. (3 points)

   b) In addition to the Gillespie simulation algorithm itself, explain the relationship between the expected value of the stochastic kinetic model (i.e., coupled chemical reactions) and the continuous deterministic formulation (i.e., ODE system). (3 points)

5. Explain the principles of Bayesian model selection (including the marginal likelihood), and also explain how the Bayesian approach can be applied for choosing an optimal biological network model. You can use any of the modeling frameworks (ODE models, ODE models with the first order approximation (=regression), Gaussian processes, Bayesian networks, etc.) as an example. (Note: exact technical details of e.g. the marginal likelihood are NOT required.) (6 points)