CS-E5885 Modeling biological networks

Exam, February 23, 2023

Lecturer: Harri Lähdesmäki

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. **Use mathematical notation and equations in your answers whenever possible and reasonable.** The use of diagrams and drawings is also encouraged.

Questions:

1. Consider random variables $\mathbf{x} = (x_1, \ldots, x_6)$.
   (a) Write the joint probability distribution for $\mathbf{x}$ in product-form in two different ways: such that if factorizes according to the undirected graphical model (also known as Markov random field) shown in Figure 1 a), and such that if factorizes according to the directed graphical model (also known as Bayesian network) shown in Figure 1 b). Explain all the mathematical notations used in your answer. (5 points)
   (b) Convert the static directed graphical model in Figure 1 b) to a dynamic Bayesian network for $\mathbf{x}(t) = (x_1(t), \ldots, x_6(t))$, $t = 0, 1, 2, \ldots$, by "unrolling" the static network over one time step, from time $t-1$ to t. Draw the resulting dynamic Bayesian network model as a directed graph for one time step, from time $t-1$ to t. Also write the joint distribution over a finite length time-series $\mathbf{x}(0), \mathbf{x}(1), \ldots, \mathbf{x}(T)$ in a product form such that the product involves conditional probabilities only for univariate variables. You can assume that the initial state $\mathbf{x}(0)$ is fixed, i.e., $\mathbf{x}(0)$ does not have any uncertainty. (5 points)

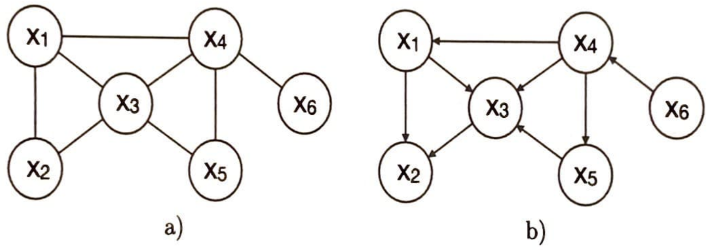

Figure 1: a) Undirected and b) directed graphical model for random variables $\mathbf{x} = (x_1, \ldots, x_6)$.

2. The reaction equations for the well-known Lotka-Volterra can be written as

$$\begin{aligned} Y_1 &\rightarrow 2Y_1 \\ Y_1 + Y_2 &\rightarrow 2Y_2 \\ Y_2 &\rightarrow \emptyset. \end{aligned}$$

   Write the ODE version of the Lotka-Volterra model for concentrations $[Y_1]$ and $[Y_2]$ assuming the mass-action kinetics with deterministic reaction rates k_1, k_2 and k_3 for the three reactions. (5 points)

3. Consider modeling continuous-time dynamics for n variables, $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))$, using ODEs, where the model for the ith variable can be written in a general form as

$$\frac{dx_i(t)}{dt} = f_i(\mathbf{x}(t)|\theta_i), \quad i = \{1, \ldots, n\}.$$

   Assume you have $N + 1$ measurements $\mathbf{y}(t_0), \mathbf{y}(t_1), \ldots, \mathbf{y}(t_N)$ from a single time-series trajectory that are measured at arbitrary time points $t_0, t_1, \ldots, t_N$. Each measurement

$\mathbf{y}(t) = (y_1(t), \ldots, y_n(t))$ contains a measurement for all n variables, where $y_i(t) = x_i(t) + \epsilon_i(t)$, where $\epsilon_i(t)$ denotes measurement error. Describe the gradient matching method for learning the parameters θ_i separately for each variable x_i. (5 points)

4. **Background:** Recall the Poisson timestep method for approximative simulation of coupled chemical reaction networks. Briefly, time is discretized into small time increments Δt. For each time increment we need to simulate a v-dimensional reaction vector r such that $r_i \sim \text{Po}(h_i(x, c_i)\Delta t)$ and then the updated state x and time t are obtained by $x := x + Sr$ and $t := t + \Delta t$, where S is the stoichiometric matrix.
Question: Describe how you can obtain the chemical Langevin equation (CLE) as a further approximation of the Poisson timestep method, and write down a general equation for CLE in terms of a stochastic differential equation (SDE) model (or the corresponding Euler-Maruyama simulation algorithm). In your answer, remember to describe the statistical aspects and motivations for deriving the CLE. (10 points total)