Continuous time Markov chains (week 10)

1 Stochastic simulation of Lotka and Volterra's predator-prey model. Lotka and Volterra's (LV) model involves a prey species X and a predator species Y. The reactions that define the model describe the birth and death of prey and predator molecules. As shorthand notation for the model we use

$$R_1 : X \qquad \xrightarrow{c_1} 2X$$
$$R_2 : X + Y \xrightarrow{c_2} 2Y$$
$$R_3 : Y \qquad \xrightarrow{c_3} \emptyset \tag{1}$$

The first reaction R_1 denotes spontaneous reproduction of prey X. Reaction R_2 denotes consumption of a prey molecule X to generate a copy of the predator molecule Y. Reaction R_3 denotes spontaneous death of predators Y. Just specifying the reactions that occur is not enough to describe the dynamic behavior of the system. We also need to specify the rate at which the reactions occur. Therefore, to complete the description we interpret X and Y as the number of molecules of each species and introduce hazard functions $h_1(X,Y)$, $h_2(X,Y)$, and $h_3(X,Y)$.

The definition of the hazard function $h_i(X,Y)$ is as follows. Given that the number of molecules are X and Y at time t, the probability of reaction R_i occurring in time interval $(t, t+s]$ is given by $1 - e^{-h_i(X,Y)s}$. Of, course, this is equivalent to saying that the time elapsed until the occurrence of reaction i is exponentially distributed with parameter $h_i(X,Y)$. It thus follows that upon defining the aggregate hazard $h(X,Y) = h_1(X,Y) + h_2(X,Y) + h_3(X,Y)$ we can model the time T until the next reaction as exponentially distributed with parameter $h(X,Y)$. Therefore, in a stochastic model of the above predator-prey relations, times T between reactions are exponentially distributed with parameter $h(X,Y)$. Once the inter-reaction time T is determined, the reaction type R_i is determined by a decision variable with probabilities proportional to the ratios $h_1(X,Y)/h(X,Y)$, $h_2(X,Y)/h(X,Y)$ and $h_3(X,Y)/h(X,Y)$. This description of the LV predator-prey model is that of a continuous time Markov chain (CTMC). To write a simulation of the system in (1) introduce the CTMC $\mathbf{X}(t) = [X(t), Y(t)]^T$ with state $\mathbf{X}(t)$ denoting the number of molecules of each type present at time t.

To complete the model we need to specify the reaction hazards $h_i(X,Y)$. Reaction R_1 happens spontaneously when a prey molecule reproduces. Assuming that prey molecules act independently it is reasonable to assume that the hazard $h_1(X,Y) = c_1 X$ is proportional to the number of prey molecules X. Similarly, Reaction R_2 happens upon the chance encounter between a prey and a predator molecule. Notice that there are XY possible pairs of prey and predator molecules. Thus, hazard $h_2(X,Y) = c_2 XY$ is reasonably modeled as proportional to the product of number of molecules X and Y. Similarly to R_1, reaction R_3 happens spontaneously, its hazard is therefore modeled as $h_3(X,Y) = c_3 Y$ proportional to the number of predator molecules Y.

Simulation of the CTMC $\mathbf{X}(t)$ with rates $h_i(\mathbf{X})$ as described above yields Gillespie's algorithm for simulation of chemical reactions. The algorithm's steps are the following

    (1)   Initialize time and CTMC's state $t = 0$, $\mathbf{X} = \mathbf{X}(0)$
    (2)   Calculate all hazards $\Rightarrow h_i(\mathbf{X})$
    (3)   Calculate transition rate $\Rightarrow \nu(\mathbf{X}) = \sum_{i=1}^{n} h_i(\mathbf{X})$
    (4)   Draw random time of next reaction $\Delta t \sim \text{Exp}(\nu(\mathbf{X}))$
    (5)   Advance time to $t = t + \Delta t$
    (6)   Draw reaction at time $t + \Delta t \Rightarrow R_i$ drawn with prob. $h_i(\mathbf{X})/\nu(\mathbf{X})$
    (7)   Update state vector $\mathbf{X}$ to account for this reaction
    (8)   Repeat from (2)

Implement Gillespie's algorithm to simulate the LV predator-prey model. Run your simulation for initial conditions $X(0) = 50$ prey and $Y(0) = 100$ predator. Set the prey reproduction rate to $c_1 = 1$ reactions/second, the rate of predator consumption of prey to $c_2 = 0.005$ reactions/second and the predator death rate to $c_3 = 0.6$ reactions/second. Run your simulation for $t = 30$ minutes. Show representative plots of your results.

2   Stochastic simulation of the lac operon While the set of genes inside a cell is immutable, proteins synthesized by them change in response to different signals in the environment. The mechanisms that control the expression of genes come in the form of auto-regulatory networks. The presence of a chemical in the environment, triggers the production of a certain protein, the production of such protein generates a change in the environment that in turns halt the production of protein. Auto-regulatory networks typically involve a group of genes, the proteins they encode for and the byproducts of the reactions of this proteins with other reactants present in the medium.

An example of an auto-regulatory gene network is the lac operon formed by three adjacent genes that control the metabolism of lactose in some bacteria. Cells can only use glucose to generate energy, but they can reduce lactose to glucose if the latter is unavailable. Such digestion of lactose into glucose occurs in presence of the enzyme β-galactosidase. A simplified model of lactose digestion and glucose consumption is therefore

$$\text{Lactose digestion:} \quad L + \beta G \xrightarrow{c_1} G + \beta G, \qquad c_1 = 1$$
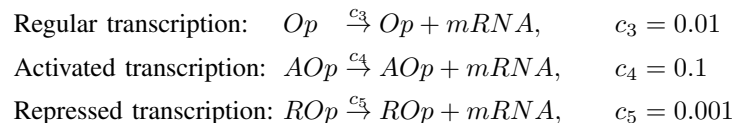$$\text{Glucose consumption:} \quad G \xrightarrow{c_2} \emptyset, \qquad c_2 = 0.1$$
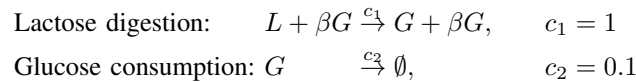
where we have use a simplified model of enzymatic reactions (compare with the enzymatic model studied in class). The first reaction models the reduction of lactose L to glucose G mediated by the enzyme βG. The second reaction models the consumption of glucose molecules G. Constant c_1 in the above expression means that lactose digestion occurs at random times exponentially distributed with parameter $c_1 L \times \beta G$ with L denoting the number of lactose molecules present and βG the number of β-galactosidase molecules present. Likewise, c_2 means that the glucose consumption occurs at exponential times with parameter $c_2 G$. Constants c_i introduced later on bear an analogous interpretation.

To enable lactose digestion as per the reactions above, cells have to produce the enzyme β-galactosidase, which in itself requires some energy expenditure. Thus, production of β-galactosidase is only justified when lactose is abundant and glucose scarce. This is controlled by the lac operon.
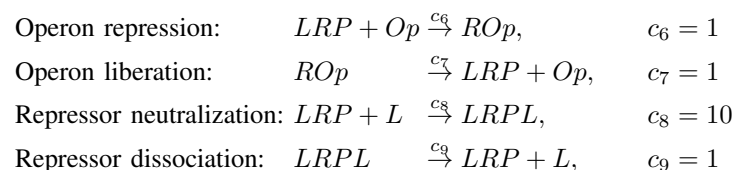
The constitutive genes on the lac operon are the promoter, the operator and the gene that encodes for β-galactosidase itself. These three genes are adjacent to each other. We note that there are in fact three different kinds of β-galactosidase, and three corresponding encoding genes, but for the discussion here we can regard them as a single gene. Promoters are integral parts of Deoxyribonucleic acid (DNA) transcription. Transcription of DNA into mRNA necessitates mediation of the enzyme RNA polymerase (RNAP). This enzyme binds to the promoter to initiate transcription of DNA into mRNA. If RNAP does not bind to the promoter, no transcription occurs. The operator is a gene located between the promoter and the lactose encoding sequence. If there is a molecule attached to the operator, it interferes with the action of the enzyme RNAP thereby halting transcription.

A model of this process is then to consider three forms of the the lac operon, the regular operon Op, the repressed operon ROp and the activated operon AOp. At different rates, transcription of DNA is possible from each of these states, whereby we have the following set of reactions
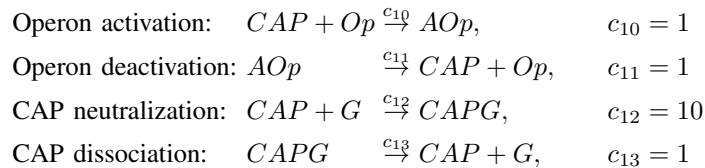
$$\text{Regular transcription:} \quad Op \xrightarrow{c_3} Op + mRNA, \qquad c_3 = 0.01$$
$$\text{Activated transcription:} \quad AOp \xrightarrow{c_4} AOp + mRNA, \qquad c_4 = 0.1$$
$$\text{Repressed transcription:} \quad ROp \xrightarrow{c_5} ROp + mRNA, \qquad c_5 = 0.001$$

The difference in the above reactions is their hazard as determined by constants c_1, c_2 and c_3. Transcription from the repressed operon ROp is much slower than transcription from the regular operon Op resulting in a much smaller hazard. Similartly, for transcription from the activated operon AOp is much larger than the hazard for transcription from the regular operon Op.

Besides promoter, operator and encoding genes, the control of β-galactosidase production involves a promoter protein and a repressor protein. Upstream of the lac operon is a gene coding for a repressor protein LRP with affinity for the operator gene. This protein is always expressed. When no lactose is present, the repressor binds to the operator thus hindering mRNA transcription and resulting in low β-galactosidase production. When lactose is present, however, the repressor binds preferentially to lactose therefore not interfering with transcription leading to increased production of β-galactosidase. The reactions modeling repression of the lac operon are then

$$\text{Operon repression:} \quad LRP + Op \xrightarrow{c_6} ROp, \qquad c_6 = 1$$
$$\text{Operon liberation:} \quad ROp \xrightarrow{c_7} LRP + Op, \qquad c_7 = 1$$
$$\text{Repressor neutralization:} \quad LRP + L \xrightarrow{c_8} LRPL, \qquad c_8 = 10$$
$$\text{Repressor dissociation:} \quad LRPL \xrightarrow{c_9} LRP + L, \qquad c_9 = 1$$
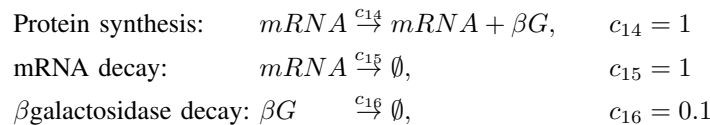
The first reaction models the adherence of the repressor protein R to the operon OP to yield a repressed operon ROp. This reaction can be undone as signified by the second reaction above. The third and fourth reactions model the binding and unbinding of repressor specimens LRP to lactose molecules L. When there is lactose present we expect the second reaction to occur more frequently than the first. This is not only because the hazard is larger but due to the much larger number of lactose proteins L with respect to the number of genes Op.

The second part of the control involves the catabolite activator protein (CAP) that when bound to the operon facilitates mRNA transcription of β-galactosidase. The amount of CAP present is inversely proportional to the amount of glucose. Hence, when glucose level decreases, the amount of CAP increases and the operon is likely to switch to its activated version. With the operon activated, the rate of mRNA transcription increases and as a consequence so does the production of β-galactosidase. The reactions to model this interaction are

$$\text{Operon activation:} \quad CAP + Op \overset{c_{10}}{\Rightarrow} AOp, \qquad c_{10} = 1$$
$$\text{Operon deactivation:} \quad AOp \overset{c_{11}}{\Rightarrow} CAP + Op, \qquad c_{11} = 1$$
$$\text{CAP neutralization:} \quad CAP + G \overset{c_{12}}{\Rightarrow} CAPG, \qquad c_{12} = 10$$
$$\text{CAP dissociation:} \quad CAPG \overset{c_{13}}{\Rightarrow} CAP + G, \qquad c_{13} = 1$$

The first and second reactions model activation and deactivation of the operon. The third and fourth reaction model binding of CAP to glucose G so that the number of free CAP molecules changes in opposite direction to the number of glucose molecules present. We note that CAP does not actually bind to glucose, but for a preliminary model the reactions above suffice.

Notice that to complete the modeling of β-galactosidase production we need to add reactions to model the synthesis of protein from mRNA molecules. These reactions are

$$\text{Protein synthesis:} \quad mRNA \overset{c_{14}}{\Rightarrow} mRNA + \beta G, \qquad c_{14} = 1$$
$$\text{mRNA decay:} \quad mRNA \overset{c_{15}}{\Rightarrow} \emptyset, \qquad c_{15} = 1$$
$$\beta\text{galactosidase decay:} \quad \beta G \overset{c_{16}}{\Rightarrow} \emptyset, \qquad c_{16} = 0.1$$

The first reaction models enzyme synthesis while the other two model degradation of $mRNA$ and βG.

When lactose and glucose are present this control mechanism results in a distinctive diauxie pattern with glucose consumed first and lactose processed after glucose is depleted.

Implement Gillespie's algorithm to simulate the lac operon. Set your initial condition to 50 glucose molecules and 50 lactose molecules. There is a single gene, originally in the regular state. There are initially 10 free LRP molecules and 10 free CAP molecules. There are no other species present at time 0. Run your simulation for the constants c_i specified above and this initial condition. Halt your simulation at $t_{\max} = 2$ hours. Constants c_i are measures in reactions/minute/molecule or reactions/minute/molecule^2.

Analyze simulation results at your digression. Questions you may want to ask yourself are, e.g., what is the average time to consume 90% of the available glucose or the time it takes to consume 90% of the lactose.