Problem Set 4

1 Degree distribution of growing networks (14 points)

COMPUTATION In this problem, we will perform simulations to explore the degree distribution of growing networks *with* and *without* preferential attachment. For the preferential attachment, we assume that the probability Π that a new node will be connected to an existing node i depends on the degree of the existing node k_i (i.e. the number of connections to other nodes in the network), in the following manner:

$$\Pi(k_i) = \frac{k_i^\alpha}{\sum_j k_j^\alpha}.$$

In the simulations, set $\alpha = 1$, $m_0 = 1$, $m = 1$, $t > 10^4$. The symbols follow the notations in the scale-free network paper that we studied in class[1]. (*Hint*: In this problem, you don't have to keep track of the entire network topology, i.e. which node is connected to which.)

a. For both networks that grow *with* and *without* preferential attachment,

   1. Plot a histogram of degrees of all the nodes, on log-linear and log-log scales.
   2. What is the mean degree? What is the degree of the most connected node?
   3. What kind of distribution do you observe?
   4. Estimate the parameters of the observed distributions.

b. Simulate and plot histograms for $\alpha = 2$, $\alpha = 0.5$. How are they different from the $\alpha = 1$ case?

2 The Feed-Forward Loop (12 points)

a. Uri Alon Exercise 4.5

b. Uri Alon Exercise 4.10

3 Discrete probability distributions (12 points)

a. The Bernoulli process is a set of N independent trials with two outcomes, one of which is called a success. The probability of success is p.

   1. Show that the distribution of the number of successes is binomial:

   $$P(k) = \binom{N}{k} p^k (1-p)^{N-k}.$$

   Find the mean and the variance of the binomial distribution.

[1] Barabasi, A-L and Albert R. Emergence of Scaling in Random Network. *Science*, 286, 509-512 (1999)

2. Show that the distribution of the number of trials to get a success is geometric:

$$P(k) = (1-p)^{k-1}p.$$

Find the mean and the variance of the geometric distribution.

b. Now consider a Poisson process, in which events occur continuously in time and independently of one another with the rate of occurrence β per unit time. After time t, the number of events occured follows a Poisson distribution:

$$P(k) = e^{-\beta t}\frac{(\beta t)^k}{k!}.$$

1. Find the mean and the variance of the Poisson distribution.
2. Let's consider the Poisson process in the context of gene expression, where each event is the transcription of an mRNA molecule. For a gene X, assume its transcription rate is 0.5 mRNA molecule per minute.
   i. What is the most likely number of mRNAs transcribed after 3 minutes? What is the mean number?
   ii. What is the probability that after 4 minutes no mRNAs were transcribed?
3. Show that the waiting time between events is exponentially distributed:

$$P(t \in (\tau, \tau + d\tau)) = \beta e^{-\beta\tau} d\tau.$$

Find the mean and the variance of the exponential distribution.

4 Distribution of the number of proteins (12 points)

a. The number of proteins per burst, k, is expected to follow a geometric distribution[2]:

$$P(k) = (1-\rho)^k \rho,$$

where $1-\rho$ corresponds to the probability that the mRNA is translated and ρ corresponds to the probability that the mRNA is degraded. Let's assume that the degradation rate of mRNA is $\gamma = 1\,min^{-1}$ and the translation rate is $\beta = 10\,min^{-1}$. What is the mean and the variance of the number of proteins produced from one mRNA?

$$mRNA \xrightarrow{\gamma} \phi$$

$$mRNA \xrightarrow{\beta} mRNA + protein$$

b. In the following, we will derive the distribution of number of proteins produced from more than one mRNA.

[2]Yu *et al.* Probing gene expression in live cells, one protein molecule at a time. Science (2006)

1. We can approximate the geometrical distribution by the exponential distribution. Show that in the limit of $\rho \ll 1$, $P(k) \approx \exp(-\rho k)\rho$. For $\rho = 0.1$, what is the probability that 1, 10 and 20 proteins are produced from one mRNA according to the exact and the approximate expression?

2. From now we change the discrete variable k to a continuous varialbe x, such that $P(k)dk = P(k) \approx \rho exp(-\rho x)dx = p(x)dx$, where $dk = 1$. $p(x) = \rho exp(-\rho x)$ is the probability density function (*pdf*) of the distribution of protein numbers x produced from a single mRNA molecule. The *pdf* of the distribution for the sum of two independent variables is the convolution of their respective *pdf*. Thus, we can find the *pdf* for the number of proteins produced from two mRNAs by convolving two identical exponential distributions. The convolution $(f * g)(x)$ of two functions $f(x)$ and $g(x)$ is defined by

$$(f * g)(x) = \int f(y)g(x-y)dy.$$

   Plot the *pdf* of: 1) an exponentially distributed random variable; 2) the sum of two independent exponentially distributed random variables.

3. Show that the probability density function for the number of proteins produced from m mRNAs follows a Gamma distribution[3] :

$$p(x) = \frac{x^{m-1} exp(-\frac{x}{b})}{(m-1)!b^m}.$$

   What is b in this equation? For which protein burst size is our continous approximation good?

[3]Equation (5) in "Linking stochastic dynamics to population distribution: an analytical framework of gene expression", by Friedman, N *et al.* PRL (2006)

8.591J / 7.81J / 7.32 Systems Biology
Fall 2014