Homework 3. Due Thursday, March 4

1. **(5 pts)**

   (a) Suppose you have n independent samples $x_1, \ldots, x_n$ of a random variable η. Prove that the estimate for the variance
   $$S_2^2 := \frac{1}{n-1}\sum_{k=1}^{n}(x_k - m)^2, \quad \text{where} \quad m = \frac{1}{n}\sum_{i=1}^{n} x_i,$$
   is unbiased.

   (b) An exponential random variable η with parameter λ has pdf $f(x) = \lambda e^{-\lambda x}$, $\lambda > 0$. Suppose that the parameter λ is unknown, but you have n independent samples $x_1, \ldots, x_n$ of η. What is the maximum likelihood estimate for λ?

2. **(5 pts)** The maximum likelihood estimation is used for finding reaction rates from molecular dynamics simulations (see e.g. [1]).

   The time interval $[t_0, t_M]$ is split into equal intervals $\Delta t = t_{m+1} - t_m$, $m = 0, 1, 2, \ldots, M-1$. The number of reactions of type j that happen in each interval $[t_m, t_{m+1}]$ is modeled as a Poisson random variable $\mathcal{P}_j$ with parameter μ_j, i.e.,
   $$P(\mathcal{P}_j = n) = \frac{\mu_j^n}{n!}e^{-\mu_j}. \tag{1}$$
   The parameter μ_j is proportional to the length of the time interval Δt, the reaction rate k_j that needs to be found, and $h_j(\mathbf{X}(t))$, a combinatorial number that can be easily calculated given the vector of concentrations of the chemical species:
   $$\mu_j = k_j h_j(\mathbf{X}(t))\Delta t. \tag{2}$$
   We assume that the time interval Δt is small enough so that $h_j(\mathbf{X}(t))$ can be approximated by a constant $h_j(\bar{\mathbf{X}}_{m+1/2})$ within it.

   We record the numbers of reactions $n_j([t_m, t_{m+1}])$ of type j occurring within each interval $[t_m, t_{m+1}]$. Given the datasets $n_j([t_m, t_{m+1}])$ and $h_j(\bar{\mathbf{X}}_{m+1/2})$, find the maximum likelihood estimate for the reaction rate k_j.

3. **(5 pts)** (This problem is based on an example from [2].) Consider the integral
   $$I = \int_0^1 \cos(x/5)e^{-5x}dx.$$
   The exact value of I is
   $$\frac{1}{626}\left(125 - 125e^{-5}\cos(1/5) + 5e^{-5}\sin(1/5)\right).$$

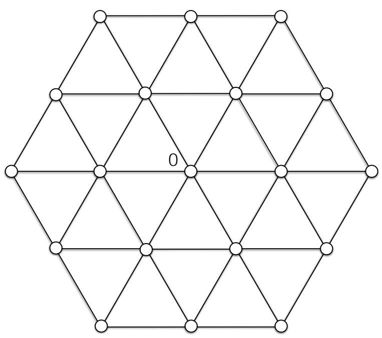
   (a) Evaluate I by Monte-Carlo (MC) as $I = E[\cos(\eta/5)e^{-5\eta}]$, where η is a random variable uniformly distributed on $[0,1]$. Make your code estimate the standard deviation $\sqrt{\mathrm{Var}(\cos(\eta/5)e^{-5\eta})}$. Estimate error of your MC result. Estimate the numbers of samples required to achieve relative errors of 1% and 0.1%.

(b) Evaluate I by Monte-Carlo using importance sampling, i.e, as $I = I_1 E[\cos(\xi/5)]$, where ξ is a random variable with the pdf

$$f_\xi(x) = \begin{cases} I_1^{-1} e^{-5x} & , x \in [0,1] \\ 0, & x \notin [0,1], \end{cases}$$

where $I_1 = \int_0^1 e^{-5x} dx$. Make your code estimate the standard deviation $\sqrt{\mathrm{Var}(\cos(\xi/5))}$. Estimate error of your MC result. Estimate the numbers of samples required to achieve relative errors of 1% and 0.1%.

4. **(5 pts)** Consider the Markov chain associated with the graph shown in the Figure below. For any vertex i, $P_{ij} > 0$ iff i and j are connected by an edge, and $P_{ij} = 1/d(i)$, where $d(i)$ is the number of edges emanating from i (the degree of i). Denote the state at the center by 0. Find the expected hitting times $k_i^{\{0\}}$ to hit 0 from each state $i \neq 0$. *Hint: using the symmetry of the problem you can dramatically decrease the number of equations in the system to be solved.*



5. **(5 pts)** Consider the discrete time Markov chain with the infinite set of states $S = \{0, 1, 2, \ldots\}$ and the transition matrix P such that $P_{i,i-1} = q$, $P_{i,i+1} = p$, $i = 1, 2, \ldots$, where $p + q = 1$, and all other entries are zero. Note that 0 is absorbing.

   (a) Assume that $p = q = 1/2$. Show that the expected hitting times $k_i^{\{0\}} = \infty$ to hit 0 for all $i \geq 1$.

   (b) Assume that $q > p$. Find the expected hitting times k_i^0 for $i \geq 1$.

References

[1] Qian Yang, Carlos A. Sing-Long, and Evan J. Reed, Learning reduced kinetic Monte Carlo models of complex chemistry from molecular dynamics, Chemical Science, Chem. Sci., 2017, 8, 5781-5796

[2] A. Chorin and O. Hald, Stochastic Tools in Mathematics and Science, 2nd edition, Springer 2009