1 Polya Urn

[5 Points]

Recall that a polya urn with $R$ initial red, and $B$ initial blue balls is a process where $R$ and $B$ are updated according to the following rule:

- Set $R_0 = R$ and $B_0 = B$
- At time $t$, draw a random ball from all $R_t + B_t$ balls, and raise the $R_t$ by one if the ball is red - otherwise raise $B_t$ by one. Denote the new number of red balls by $R_{t+1}$ and the new number of blue balls by $B_{t+1}$.

Set $X_t = 1$ if the ball drawn at time $t$ is red, and $X_t = 0$ otherwise. Fix a set $I_R \subset \{1, \ldots, n\}$ and its complement, $I_B$. We have seen that the probability of drawing a sequence $X_1, \ldots, X_n$ such that $X_i = 1$ for $i \in I_R$ and $X_i = 0$ for $i \in I_B$ does only depend on the sizes $n_1$ and $n_2$ of $I_1$ and $I_2$ and is equal to

$$\Pr(X_i = 1 \text{ for } i \in I_R \text{ and } X_i = 0 \text{ for } i \in I_B) = \frac{R(R+1) \ldots (R+n_1-1) \times B(B+1) \ldots (B+n_2-1)}{(B+R)(B+R+1) \ldots (B+R+n-1)}.$$ 

In this exercise, you will prove the following theorem

**Theorem 1** The probability of the sequence $X_1, \ldots, X_n$ draws from the Poly Urn with initially $R$ red and $B$ blue can equivalently be calculated by first drawing $p \sim \beta(R,B)$ and then choosing $X_i \ iid$ with distribution $\text{Be}(p)$.

Here $\beta(R,B)$ is the probability distribution on $[0,1]$ that has the probability density function $\frac{1}{Z} x^{R-1} (1-x)^{B-1}$, where $Z = \int_0^1 x^{R-1} (1-x)^{B-1} dx$.

1. Let $A$ and $B$ non-negative integers. Use integration by parts to calculate the integral

$$\int_0^1 x^A (1-x)^B dx.$$

2. Let $R, B, n_1$ and $n_2$ be non-negative integers, and assume that $X \sim \beta(A,B)$. Calculate first the normalization factor $Z = \int_0^1 x^{R-1} (1-x)^{B-1}$ for the $\beta$-distribution, and then the expectation of $X^{n_1} (1-X)^{n_2}$.

3. Use the result from 2 to prove the theorem.
2 Degrees for preferential attachment

Let $D_n$ be the degree of a random vertex in the preferential attachment graph on $n$ nodes (note that it has two sources of randomness - the randomness from choosing a vertex, and that of the preferential attachment graph). For the various version of preferential attachment discussed in the lectures (independent, conditional, and sequential), the random variable $D_n$ converges in distribution to a random variable $D$ whose distribution is given by

$$\Pr(D = k) = \frac{2m(m+1)}{k(k+1)(k+2)}.$$

A different way of obtaining the distribution of $D$ proceeds by proving that Preferential attachment has a weak local limit, and then establishing the degree distribution for the root of the limiting rooted graph. In that approach, the variable $D$ appears naturally as a sum of $m$ and a mixed Poisson random variable $q$:

$$D = m + q \quad \text{where} \quad q \sim \text{Poi}(\gamma \frac{1-x}{x})$$

where $\gamma \sim \Gamma(m, 1)$ and $x = \sqrt{u}$ with $u$ being a uniform random variable in $[0, 1]$. In this exercise, you will prove that the two are indeed equivalent.

1. Let $A$ be a non-negative integer. Use integration by parts to calculate the integral

$$\int_0^\infty x^A e^{-x} dx$$

2. Let $A$ and $k$ be non-negative integers, and let $c > -1$. A random variable $X$ has distribution $\Gamma(A, 1)$ if its probability density function is equal to $\frac{1}{Z} x^{A-1} e^{-x}$, where $Z$ is a normalization constant. Calculate first $Z$, and then the expectation of $X^k e^{-cX}$.

3. Calculate the probability that the random variable $q$ above takes the value $k$, conditioned on $\gamma$ and $x$.

4. By first taking expectation with respect to the random variable $\gamma$, and then with respect to $u = x^2$, get the probability that $q = k$.

5. Relate this result to the explicit formula given above and show that the two give the same probability distribution.
3 Infection digraph

[6 Points]

Recall that the exponential distribution with rate $\gamma$ is the probability distribution on $\mathbb{R}_+$ with density $\rho(T) = \gamma e^{-\gamma T}$.

For a graph $G$, recovery rate $\gamma$ and infection rate $\tau$, the infection digraph of an SIR model on $G$ is determined as follows: for all vertices $i$, draw a recovery clock $T_i \sim \exp(\gamma)$ and for all edges $\{i, j\}$ in $E(G)$ draw two infection clocks $T_{ij} \sim \exp(\tau)$ and $T_{ji} \sim \exp(\tau)$, all independently of each other. The infection digraph $D_{SIR}$ is then defined by putting an oriented edge $ij$ from $i$ to $j$ whenever $\{i, j\}$ is an edge in $G$ and $T_{ij} \leq T_i$. We set $X_{ij} = 1$ if this is the case, and $X_{ij} = 0$ otherwise. Note that the random variables $X_{ij}$ are independent if we condition on the recovery times $\{T_i\}_{i \in V(G)}$.

1. Let $p_T = 1 - e^{-\tau T}$. Use elementary calculations involving the exponential distribution $\exp(\tau)$ to show that conditioned on the recovery times, $X_{ij} \sim \text{Be}(p_T)$ whenever $\{i, j\}$ is an edge in $G$.

2. For a vertex $i$, let $N(i)$ be the set of neighbors of $i$ in $G$. Show that for a vertex $i$ of degree $d_i$ in $G$, the in-degree in $D_{SIR}$, $d_i^- = \sum_{j \in N(i)} X_{ji}$, has distribution $\text{Bin}(d_i, p)$, where $p = \frac{\tau}{\gamma + \tau}$. Write down the generating function $G_i^-(\eta)$ for the random variable $d_i^-$. 

3. Show that conditioned on $T_i$, the out-degree of $i$ in $D_{SIR}$, $d_i^+ = \sum_{j \in N(i)} X_{ij}$, has distribution $\text{Bin}(d_i, p_T)$. Write down the generating function $G_i^+(\eta|T_i)$ for the random variable $d_i^+$ conditioned on $T_i$.

Note that the generating function of the unconditioned random variable $d_i^+$ is just the expectation of $G_i^+(\eta|T_i)$ with respect to $T_i$,

$$G_i^+(\eta) = \gamma \int G_i^+(\eta|T)e^{-\gamma T}dT.$$ 

(this is just an observation, you don’t need to simplify this further for the moment).

4. Specializing to $G = K_n$ and $\tau = \beta/n$, calculate $G^-(\eta)$ and $G^+(\eta|T_i)$ in the limit $n \to \infty$ (if you don’t make a mistake, you should get the generating functions of two Poisson random variables; also, since $K_n$ is regular, the result will not depend on $i$, so we dropped the index $i$). What are the rates $e^-$ and $e^+(T_i)$ of these two Poisson random variables?

5. Using the above, it is easy to calculate the generating function $G^+$ of the unconditioned out-degree $d_i^+$, since by dominated convergence it converges to the expectation of the limit you just calculated. Calculate this expectation, and express the limiting generating function of $d_i^+$ as an explicit function of $R_0 = \beta/\gamma$ and $\eta$. Compare the expression you get to that of a geometric random variable.

6. (Bonus) Explicitly solve the equation $\theta + G^+(1 - \theta) = 1$ for the survival probability $\theta$ of a birth process with off spring distribution $d^+$ (in the limit $n \to \infty$, where, as discussed in class, it is the survival probability of the infection). Compare the results to those obtain from $G^-$, both for $R_0 = 1 + \epsilon$ for small $\epsilon$ (in which case I’d like you to compare the rate at which $\theta \to 0$ as $\epsilon \to 0$) as well as for large $R_0$ (in which case I’d like you to compare the rate at which $\eta = 1 - \theta \to 0$ as $R_0 \to \infty$).