Rumours, Consensus and Epidemics on Networks
Problem Sheets

1 Preliminaries

1. Let X_1 and X_2 be independent Poisson random variables with means λ_1 and λ_2. Using generating functions, show that $X_1 + X_2$ is Poisson with mean $\lambda_1 + \lambda_2$.

2. Let T have an Exponential distribution with parameter μ.

   (a) Show that the distribution is "memoryless", i.e., show that for all $t, u > 0$,
   $$\mathbb{P}(T > t + u | T > u) = \mathbb{P}(T > t).$$

   (b) If $c > 0$ is a given constant, then show that the random variable $\tilde{T} = cT$ has an Exponential distribution with parameter μ/c.

   (c) Let T_1 and T_2 be independent Exponential random variables with parameters λ_1 and λ_2 respectively, and let $T = \min\{T_1, T_2\}$.

      i. Show that the distribution of T is Exp($\lambda_1 + \lambda_2$).

      ii. Show that the probability that $T = T_1$ is $\lambda_1/(\lambda_1 + \lambda_2)$, irrespective of the value of T.

3. Let $X_t, t \geq 0$ be a Poisson process of rate λ, and let $c > 0$ be a given constant. Define $Y_t = X_{ct}$ for all $t \geq 0$. Show that $Y_t, t \geq 0$ is a Poisson process of rate $c\lambda$.

4. Let $N_t^1, t \geq 0$ and $N_t^2, t \geq 0$ be independent Poisson processes of rate λ_1 and λ_2 respectively. Let $N_t = N_t^1 + N_t^2$ denote their superposition.

   (a) Use the answer to Question 1 to show that $N_t, t \geq 0$ is a Poisson process of rate $\lambda = \lambda_1 + \lambda_2$.

   (b) Use the answer to Question 2 to show the same thing.

5. We say that a random variable N has a Geometric distribution with parameter p, written $N \sim Geom(p)$ if
   $$P(N = k) = p(1-p)^{k-1}, \quad k = 1, 2, 3, \ldots$$
   Let $N \sim Geom(p)$, and let $T_1, T_2, T_3, \ldots$ be iid Exp(λ) random variables, independent of N. Let $T = \sum_{k=1}^{N} T_k$. Using moment generating functions or otherwise, show that T is exponentially distributed with parameter λp. (***Hint.*** Recall that the moment generating function of T is defined as $M(\theta) = \mathbb{E}[\exp(\theta T)]$. First compute $\mathbb{E}[\exp(\theta T) | N = n]$ and then average over N to obtain the unconditional expectation.)

6. Let $X_t, t \geq 0$ be a Poisson process of rate λ, and let $Y_1, Y_2, Y_3, \ldots$ be iid Bernoulli(p) random variables. Recall that this means that $Y_i = 1$ with probability p and $Y_i = 0$ with probability $1 - p$.

   Let $X_t^1 = \sum_{i=1}^{X_t} Y_i$ be the process obtained by retaining each point of the Poisson process X_t independently with probability p and discarding it with probability $1 - p$. It is called the Bernoulli(p) thinning of the Poisson process X_t.

   Using the answer to question 5, show that $X_t^1, t \geq 0$ is a Poisson process of rate λp by showing that the times between successive events are Exp(λp).

2 Rumour Spreading

1. Let $G = (V, E)$ be the complete, undirected graph on n nodes. The simple epidemic on G is described as follows. There is an initial set $S \subseteq V$ of infected nodes, and all other nodes are healthy. Each infected node attempts to spread infection at the points of a unit rate Poisson process. The Poisson processes corresponding to distinct nodes are mutually independent.

   When a node attempts to spread infection, it chooses a node from V uniformly at random (including itself), independent of the Poisson processes of spreading times and past node choices of itself or of other nodes. If the chosen node is healthy, it becomes infected at this time. If it is already infected, nothing changes. Once a node becomes infected, it remains infected forever.

   In answering the questions below, think of n as large. You may replace sums by integrals, ignore terms of smaller order than the dominant term, and make any other reasonable approximations required, so long as you get the correct dominant term as a function of n.

   (a) Suppose a single node is initially infected. Compute the mean time until at least $\sqrt{n}$ nodes are infected. Call this $\mathbb{E}[T(\sqrt{n})]$.

   (b) What is the mean time until at least $n/2$ nodes are infected? Call this $\mathbb{E}[T(n/2)]$. What is the smallest constant c such that $\mathbb{E}[T(n/2)] \leq c\mathbb{E}[T(\sqrt{n})]$ for all n sufficiently large?
   Hint. Replacing sums by integrals and making other reasonable approximations may help you estimate c.

   (c) Next, suppose that $\sqrt{n}$ nodes are initially infected. (Assume that n is a perfect square.) What is the mean time until all n nodes are infected? Call this $\mathbb{E}[\tilde{T}(n)]$. What is the smallest constant c such that $\mathbb{E}[T(n/2)] \leq c\mathbb{E}[\tilde{T}(n)]$ for all n sufficiently large?

2. Consider the following model of the spread of two competing rumours on K_n, a complete graph on n nodes. We will call them the red and blue rumour and can think of them as fake news and true news spreading on the same network. Initially, each rumour is known to a single node, different for the two rumours. Each node will adopt/ believe whichever rumour it hears first and will only spread that rumour. If it subsequently hears the other rumour, it will ignore it.

There are independent unit rate Poisson processes at each node. Whenever there is an increment of the Poisson process $N_v(\cdot)$ at node v, this node will spread the rumour it has adopted (if it has adopted one by this time) to another node chosen uniformly at random from the population. The recipient will adopt the rumour if it has not yet adopted one, and will ignore it otherwise. The process continues until all nodes have adopted a rumour.

The fraction of nodes which have adopted each rumour at the end are (dependent) random variables, which sum to 1. Let p_R denote the fraction that have adopted the red rumour. What can you say about the random variable p_R if n is large?

3. Consider the following rumour spreading process. There is a population of n individuals, one of whom initially knows a rumour. There are n independent unit rate Poisson processes, one associated with each individual. If individual i knows the rumour at an increment time of his Poisson process, then at this time he picks an individual uniformly at random from the population (including himself!), and informs this person of the rumour. This counts as one communication, even if the target (which could be himself) already knows the rumour. The choice of individual to contact at each step is independent of all past choices.

   How many communications does it take, on average, until all n individuals know the rumour? Obtain a simple approximate expression for this quantity, for large n; you may replace sums by integrals, or make other reasonable approximations, without justifying them.

4. In this problem, we prove Theorem 2 in Section 2 of the lecture notes. Let T_k denote the first time that exactly k nodes know the rumour, and define $X_k = T_{k+1} - T_k$. We saw that the X_k are mutually independent, and that X_k is exponentially distributed with parameter $k(n-k)/n$. Let $a_k = \max\{k, n-k\}/n$ and let $Y_k = a_k X_k$.

   (a) Show that $\{Y_k, k = 1, \dots, n\}$ are independent and exponentially distributed random variables with parameters $\min\{k, n-k\}$, defined on the same probability space as the X_k. In particular, if $k \le n/2$, then $a_k = (n-k)/n$ and $Y_k \sim Exp(k)$.

   (b) Now consider the random variable
   $$T_{\lfloor n/2 \rfloor + 1} := \sum_{k=1}^{\lfloor n/2 \rfloor} T_{k+1} - T_k = \sum_{k=1}^{\lfloor n/2 \rfloor} Y_k + Z_k \text{ where } Z_k = X_k - Y_k.$$
   Show that
   $$\sum_{k=1}^{\lfloor n/2 \rfloor} Z_k \xrightarrow{p} \log 2 \text{ as } n \to \infty.$$

   (c) Show that $\sum_{k=1}^{\lfloor n/2 \rfloor} Y_k$ has the same distribution as $\max_{k=1}^{\lfloor n/2 \rfloor} V_k$, where $V_1, V_2, \dots$ are iid $Exp(1)$ random variables.
   Hint. Let $V^{(1)} \le V^{(2)} \le \dots \le V^{(\lfloor n/2 \rfloor)}$ denote the order statistics of the V_i. What can you say about the distribution of $V^{(1)}$? And that of $V^{(2)} - V^{(1)}$, conditional on $V^{(1)}$? And that of $V^{(3)} - V^{(2)}$ conditional on $V^{(1)}$ and $V^{(2)}$?

   (d) Show by direct calculation that $\max\{V_1, V_2, \dots, V_{\lfloor n/2 \rfloor}\} - \log(\lfloor n/2 \rfloor) \xrightarrow{d} \zeta_1$, where ζ_1 has a standard Gumbel distribution.

(e) Put together the answers to earlier parts to show that $T_{\lfloor n/2 \rfloor+1} - \log n \xrightarrow{d} \zeta_1$.
Hint. Look up Slutsky's theorem.

(f) Notice that $T_{\lfloor n/2 \rfloor+1}$ has the same distribution as $T_n - T_{\lfloor n/2 \rfloor+1}$ and is independent of it. (More precisely, this is true if n is odd. If n is even, we need to deal more carefully with the term $T_{\lfloor n/2 \rfloor+1} - T_{\lfloor n/2 \rfloor}$, but this is a minor annoyance; this random variable tends to zero in probability and can be dealt with by using Slutsky's theorem once more.) Use this fact to complete the proof of Theorem 2. You may take for granted that, if A_n and B_n are independent random sequences such that $A_n \xrightarrow{d} A$ and $B_n \xrightarrow{d} B$, then $(A_n, B_n) \xrightarrow{d} (A, B)$.
Hint. Look up the Continuous Mapping Theorem.

5. Prove Theorem 3 in Section 2.

6. Let S_n be the star graph on n nodes consisting of a single hub node connected to each of $n-1$ leaves; there are no edges between leaves. Consider the following rumour-spreading model. Each node v becomes active at the increment times of a Poisson process of rate $\lambda_v = 1$. When a node v becomes active, it picks a node w with probability $p_{vw} = 1/n$ if v is the hub and w is a leaf, and with probability $p_{vw} = 1$ if v is a leaf and w is the hub; $p_{vw} = 0$ if v and w are both leaves.

   (a) Suppose that only the hub node knows the rumour at time 0. Compute $\mathbb{E}[T_{k+1} - T_k]$ exactly, and use this to compute $\mathbb{E}[T_n]$ exactly.

   (b) Compute the matrix R of spreading rates and its conductance, $\Phi(R)$, defined in Section 2 Compute the corresponding upper bound on $\mathbb{E}[T_n]$, and compare it with the exact answer.

   (c) Repeat the exact analysis when only a single leaf node initially knows the rumour.

7. Let C_n be the cycle graph on n nodes numbered $\{1, 2, 3, \ldots, n\}$, where there are two directed edges out of each node i. These go to nodes $i-1$ and $i+1$ for $2 \leq i \leq n-1$. The edges out of node 1 go to nodes 2 and n, while the edges out of node n go to nodes $n-1$ and 1. Consider the rumour-spreading model of the previous question, with $\lambda_v = 1$ for all v and $p_{ij} = 1/2$ if for every $(i, j) \in E$.

   (a) Suppose that only a single node knows the rumour at time 0. Compute $\mathbb{E}[T_{k+1} - T_k]$ exactly, and use this to compute $\mathbb{E}[T_n]$ exactly. (*Hint.* Observe that, at any time, the set of nodes that knows the rumour has to be a contiguous set.)

   (b) Compute R and $\Phi(R)$ for the cycle graph with the rates and probabilities specified above, and the corresponding upper bound on $\mathbb{E}[T_n]$, and compare it with the exact answer.

3 Averaging and Consensus

1. Let $G = (V, E)$ be an undirected graph. Initially, each node $v \in V$ is assigned a value $x_v(0)$ in the interval $[0, 1]$. Time is discrete, and nodes update their values synchronously according to the linear recursion

$$x_v(t+1) = \frac{1}{\deg(v)} \sum_{u:(u,v)\in E} x_u(t). \tag{1}$$

   (a) Write down the set of linear equations in (1) in matrix form as $\mathbf{x}(t+1) = P\mathbf{x}(t)$, i.e., specify the elements of the matrix P.

   (b) Compute an invariant distribution corresponding to the stochastic matrix P, i.e., find a solution of $\pi P = \pi$.
   Hint: It turns out the Markov chain with transition probability matrix P is reversible, and you can compute an invariant distribution by solving the local balance equations, which state that

$$\pi_x p_{xy} = \pi_y p_{yx} \quad \forall\, x, y \in V.$$

   If these equations have a probability vector π as a solution, then it is an invariant distribution of the Markov chain.

   (c) Assume that the graph G is connected and non-bipartite. (A graph is bipartite if the vertex set V can be partitioned into disjoint subsets V_1 and V_2, i.e., with $V_1 \cup V_2 = V$ and $V_1 \cap V_2 = \emptyset$, such that there are no edges between two nodes of V_1 or two nodes of V_2. In other words, $E \subseteq V_1 \times V_2$.) In this case, it is known that the Markov chain with transition probability matrix P is irreducible and aperiodic. Comment on what happens to $\mathbf{x}(t)$ as t tends to infinity.

   (d) What property of a node determines how influential that node is in determining the final outcome of the above process?

2. The Wright-Fisher model is a discrete time model describing the evolution of a population of N genes. Each gene has two forms, or alleles, which we denote A and a. The population size stays fixed over time. If we let N_t denote the number of A alleles in generation t, then generation $t+1$ is obtained as follows. Each of the N genes in generation $t+1$ is sampled independently, and uniformly at random (with replacement), from the genes in generation t.

   (a) Conditional on $N_t = n$, the number of A alleles in generation $t+1$ has a binomial distribution. What are the parameters of this binomial distribution?

   (b) Use the answer to the last part to show that N_t is a martingale.

   (c) If $N_0 = k$, what is the probability that eventually there are only A alleles in the population? Explain your answer carefully.

3. Let $G = (V, E)$ be a graph on 4 nodes with the following 5 edges $(1, 2)$, $(2, 3)$, $(3, 4)$, $(4, 1)$ and $(1, 3)$; in other words, it is a square with one diagonal. Think of each of these edges as two directed edges. Now consider the voter model on this graph where contacts along

each directed edge happen according to independent unit rate Poisson processes. Suppose the voter model starts with nodes 1 and 3 in state 1, and nodes 2 and 4 in state 0. Compute the probability that all nodes eventually reach consensus on the value 1.
Hint. If you need to find an invariant distribution, see if the local balance equations have a solution.

4. Let S_n denote the star graph, which consists of a hub connected to each of $n-1$ leaves; there are no edges between leaves. Consider the following voter model on S_n. Each node becomes active at the points of a Poisson process of rate 1, independent of all other nodes. When it becomes active, it chooses a neighbour uniformly at random from the set of all its neighbours (i.e., excluding itself), and copies the state of that neighbour.

   Denote by $X_v(t) \in \{0,1\}$ the state of node v at time t. Let $M(t) = (n-1)X_{\text{hub}}(t) + \sum_{v \neq \text{hub}} X_v(t)$.

   (a) Show that $M(t)$ is a martingale.

   (b) Suppose that initially the hub and $k-1$ leaves are in state 1, while $n-k$ leaves are in state 0. What is the probability of being absorbed into the all-1 state?

5. In this problem, we compute an upper bound on the time to reach consensus for the voter model on a star graph described in the previous problem.

   (a) We would like to describe the voter model, backwards in time, in terms of coalescing random walks. For a single one of these random walks, what are the transition rates (from a leaf to the hub, and from the hub to each leaf)?

   (b) Next, let us consider two of these random walks, started at different leaves, say. We only need to keep track of the distance between the particles performing these random walks. This distance is either 0, 1 or 2, and when it becomes 0, the particles are at the same node and coalesce. Describe the evolution of this distance as a Markov process.

   (c) Compute the expected time for this Markov process to hit state (distance) 0 starting in state 2. This is the expected time for two random walks to meet, and hence for two particles to merge.

   (d) We are interested in the time until each of $n-1$ other particles has merged with a given particle. This is an upper bound on the time to consensus. Using the fact that the expectation of the maximum of non-negative random variables is bounded by the expectation of their sum, obtain an upper bound on the expected time to reach consensus in the voter model on the star.

6. Consider n nodes arranged in a ring, with an edge between each node and its two neighbours. In other words, if the nodes are numbered $0, 1, 2, \ldots, n-1$, node i has edges to nodes $i-1$ and $i+1$ modulo n. Consider the voter model on this graph, where each node becomes active at the points of a unit rate Poisson process (independent of other nodes), chooses one of its two neighbours with equal probability (independent of everything else), and adopts the state of that neighbour. We want to bound the time to consensus in this model.

(a) Consider two particles, starting at nodes i and j, and performing independent random walks until they meet. Let Y_t denote the clockwise distance from i to j; suppose that initially j lies clockwise of i so that $Y_0 \in \{1, 2, \ldots, n-1\}$. The two particles merge when this distance becomes 0 or n, i.e., at the time $T = \inf\{t > 0 : Y_t = 0 \text{ or n}\}$.

Show that $M_t = Y_t^2 - 2t$ is a martingale on the time period $[0, T]$.

(b) Compute $\mathbb{E}T$, the expected time until the two particles coalesce. You may use the fact that

$$\mathbb{P}(Y_T = n) = \frac{Y_0}{n}, \quad \mathbb{P}(Y_T = 0) = 1 - \frac{Y_0}{n}.$$

(c) We are interested in the time until each of $n-1$ other particles has merged with a given particle. This is an upper bound on the time to consensus. Using the fact that the expectation of the maximum of non-negative random variables is bounded by the expectation of their sum, obtain an upper bound on the expected time to reach consensus in the voter model on the cycle graph on n nodes.