UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Homework 10
Spring 2023

1. **Midterm**

   Solve again the midterm problems which you got incorrect, resorting to the solutions if necessary. Make sure that you understand the concepts tested without simply copying the solutions, as you may find similar questions on the final exam.

   Solution: See midterm solutions.

2. **Connected Random Graph**

   We start with the empty graph on n vertices. Iteratively, we add an undirected edge, chosen uniformly at random from the edges that are not yet present in the graph, until the graph is connected.

   Hint: Recall the coupon collector's problem.

   a. Suppose that there are currently k connected components in the graph. Let X_k be the number of edges we need to add until there are $k-1$ connected components. Show that $\mathbb{E}(X_k) \leq \frac{n-1}{k-1}$.
   b. Let X be the total number of edges in the final connected graph. Show that $\mathbb{E}(X) \leq Cn \log n$ for some constant C.

   Solution:

   a. Let p_k be the probability that the first edge we add brings us to $k-1$ components. As we continue to add edges, the probability that each new edge added will produce $k-1$ components from k components is increasing, starting from p_k. If $Y_k \sim \text{Geometric}(p_k)$, then $\mathbb{E}(X_k) \leq \mathbb{E}(Y_k) = \frac{1}{p_k}$. (*See remark.*)
   To bound p_k, suppose there are currently k components and u is one endpoint of the edge that we are currently adding. There are at most $n-1$ vertices to which we can connect u, and at least $k-1$ of these will reduce the number of components, so $p_k \geq \frac{k-1}{n-1}$.
   b. Observe that $X = \sum_{k=2}^{n} X_k$. Using part a,

$$\mathbb{E}(X) = \sum_{k=2}^{n} \mathbb{E}(X_k) \leq \sum_{k=2}^{n} \frac{n-1}{k-1} = (n-1)H_{n-1} \leq n \log n.$$

Remark. Y_k is intuitively "larger" than X_k, but it is difficult to explain the precise meaning of this in the context of randomness. We say that Y_k *stochastically dominates* X_k if $\mathbb{P}(Y_k \geq x) \geq \mathbb{P}(X_k \geq x)$ for all x, which holds here and implies that $\mathbb{E}(Y_k) \geq \mathbb{E}(X_k)$.

To explain why, we could use the *tail sum* formula or consider a **coupling** argument: suppose that each time we add an edge, we flip a coin of probability p_k. If the coin comes up heads, we add an edge that connects two components; otherwise, we still have some additional chance of connecting two components. In this case, the number of edges until we have $k-1$ components is at most the number of flips until we see heads, or $X_k \leq Y_k$, so $\mathbb{E}(X_k) \leq \mathbb{E}(Y_k)$.

3. **Isolated Vertices**

   Consider an Erdős–Rényi random graph $\mathcal{G}(n, p(n))$, where n is the number of vertices and $p(n)$ is the probability that any specific edge appears in the graph. Let X_n be the number of isolated vertices in $\mathcal{G}(n, p(n))$.

   a. Show that $\mathbb{E}(X_n) \to \exp(-c)$ as $n \to \infty$ when $p(n) = \frac{(\ln n)+c}{n}$ for some constant c.
   b. Conclude that $\mathbb{E}(X_n) \to \infty$ when $p(n) \ll \frac{\ln n}{n}$.
   c. Conclude that $\mathbb{E}(X_n) \to 0$, and $X_n \to 0$ in probability, when $p(n) \gg \frac{\ln n}{n}$.

   The asymptotic notation $f(n) \ll g(n)$ means that $\frac{f(n)}{g(n)} \to 0$ as $n \to \infty$.

   Hint: From Taylor series expansion, $\ln(1+x) \approx x$ when x is small.

   Solution:

   a. The probability that any specific vertex is isolated is $(1-p(n))^{n-1}$, so

   $$\mathbb{E}(X_n) = n(1-p(n))^{n-1}.$$

   When $p(n) = \frac{(\ln n)+c}{n}$, using the approximation $\ln(1+x) \approx x$ for small x,

   $$\ln \mathbb{E}(X_n) = \ln n + (n-1)\ln\left(1 - \frac{(\ln n)+c}{n}\right) \sim \ln n - \frac{(n-1)((\ln n)+c)}{n} \to -c.$$

   Thus $\mathbb{E}(X_n) \to \exp(-c)$.

   b. When $p(n) \ll \frac{\ln n}{n}$, we have $p(n) < \frac{(\ln n)+c}{n}$ for all c, which means that the limit of $\mathbb{E}(X_n)$ is lower bounded by $\exp(-c)$ for all c, or $\mathbb{E}(X_n) \to \infty$.

   c. When $p(n) \gg \frac{\ln n}{n}$, then by the same reasoning, the limit of $\mathbb{E}(X_n)$ is upper bounded by $\exp(-c)$ for all c, or $\mathbb{E}(X_n) \to 0$. Now, by Markov's inequality,

   $$\mathbb{P}(X_n > 0) = \mathbb{P}(X_n \geq 1) \leq \mathbb{E}(X_n) \to 0,$$

   so $X_n \to 0$ in probability as $n \to \infty$.

   Remark. We have shown that $\frac{\ln n}{n}$ is a *threshold* for the expected number of isolated vertices, but it is also a threshold for connectivity: if $p(n) = \lambda\frac{\ln n}{n}$, then the probability that the graph is connected tends to 1 when $\lambda > 1$ and tends to 0 when $\lambda < 1$.

   $p(n) = \lambda\frac{\ln n}{n}$ is called a *coarse* parameterization and $p(n) = \frac{(\ln n)+c}{n}$ a *fine* parameterization. If a finer parameterization is used, then we can observe subtler, "smoother" transitions instead of "sharp" thresholds. In fact, it is known that for $p(n) = \frac{(\ln n)+c}{n}$, $X_n \to \text{Poisson}(\exp(-c))$ in distribution as $n \to \infty$.

4. **Random Bipartite Graph**

Consider a random bipartite graph with K left nodes and M right nodes. Each of the $K \cdot M$ possible edges of this graph is present independently with probability p.

a. Find the distribution of the degree of a particular right node.

b. Now fix a left node u and right node v. Conditioned on the event that the edge (u, v) is present, find the distribution of the degree of v. Is what you find the same as in part a?

c. Call a right node with degree one a *singleton*. What is the expected number of singletons in a random bipartite graph?

d. Find the expected number of left nodes connected to at least one singleton.

Solution:

a. The number of edges connected to a right node has Binomial(K, p) distribution, as each left node is connected to it with probability p, independently of all other left nodes, so

$$\mathbb{P}(\deg(\text{right node}) = d) = \binom{K}{d} p^d (1-p)^{K-d}.$$

b. Given that an edge is already present, the degree of v is distributed as $1+$Binomial$(K-1, p)$ by the same reasoning as in part a, so

$$\mathbb{P}(\deg(v) = d \mid (u, v) \text{ is an edge}) = \binom{K-1}{d-1} p^{d-1} (1-p)^{K-d}$$

where $1 \leq d \leq K$, as opposed to $0 \leq d \leq K$ in part a.

c. By linearity of expectation, the expected number of singletons is

$$M \cdot \mathbb{P}(\text{a right node is singleton}) = M \binom{K}{1} p^1 (1-p)^{K-1} = MKp(1-p)^{K-1}.$$

d. We first find the probability of a left node ℓ being connected to at least one singleton:

$$\begin{aligned}
&1 - \mathbb{P}(\ell \text{ is not connected to any singleton}) \\
&= 1 - \prod_{i=1}^{M} (1 - \mathbb{P}(\ell \text{ is connected to } i\text{th right node and } i \text{ is singleton})) \\
&= 1 - (1 - p(1-p)^{K-1})^M.
\end{aligned}$$

Note that the connections are independent. Then, by linearity of expectation, the average number of left nodes that are connected to at least one singleton is

$$K \cdot (1 - (1 - p(1-p)^{K-1})^M).$$

Alternate solution. We still first find the probability that a left node ℓ is connected to at least one singleton. Let us condition on $\deg(\ell) = d$, observing that the degree of ℓ has Binomial(M, p) distribution. Considering one of the d edges,

$$\mathbb{P}(\text{right node connected to this edge is a singleton}) = (1-p)^{K-1}.$$

Thus, the probability that none of the d right nodes connected to ℓ are singletons is

$$\mathbb{P}(\ell \text{ not connected to any singletons} \mid \deg(\ell) = d) = (1 - (1-p)^{K-1})^d.$$

By the law of total probability, conditioning on the degree of the left node, the probability that a left node is connected to at least one singleton is

$$\begin{aligned}
&\sum_{d=1}^{M} (1 - (1 - (1-p)^{K-1})^d) \cdot \binom{M}{d} p^d (1-p)^{M-d} \\
&= \sum_{d=0}^{M} \binom{M}{d} p^d (1-p)^{M-d} - \sum_{d=0}^{M} \binom{M}{d} (p - p(1-p)^{K-1})^d (1-p)^{M-d} \\
&= 1 - (1 - p(1-p)^{K-1})^M.
\end{aligned}$$

We can then use linearity of expectation as before.

5. **Random Graph with Partition**

Let $G = (V, E)$ be a graph on n vertices, where the vertex set V is deterministically partitioned into two disjoint subsets V_1 and V_2. You do not know the sizes of V_1 and V_2, but assume they both contain at least $n/100$ vertices.

The edges of G are independently randomly generated as follows. If u and v belong to different subsets, then the edge $\{u, v\}$ appears with probability 1. Otherwise, the edge $\{u, v\}$ appears with probability $p < 1$, where p does not depend on n.

a. Is G an Erdős–Rényi random graph?

b. Suppose you do not know which vertices are in V_1 or V_2. You try to determine this by finding a partition $V = \hat{V}_1 \sqcup \hat{V}_2$ such that $\hat{V}_1, \hat{V}_2$ are nonempty subsets and $\{u, v\} \in E$ for all $u \in \hat{V}_1, v \in \hat{V}_2$. If there are multiple candidates, you choose one arbitrarily.

   Your choice is considered correct if $\hat{V}_1 = V_1$ or $\hat{V}_1 = V_2$ (since both identify the same partition of V). What is the limit as $n \to \infty$ of the probability that the procedure above recovers the correct partition?

   Hint. Consider the complementary graph G', in which an edge appears if and only if it does not appear in the original graph G.

Solution:

a. No, as the edges in G are not generated with identical probabilities.

b. In the complementary graph G', no edges cross between V_1 and V_2, and other edges are generated independently with identical probability $q = 1 - p > 0$. In other words, G' is the disjoint union of two Erdős–Rényi random graphs $G_1 \sim \mathcal{G}(|V_1|, q)$ and $G_2 \sim \mathcal{G}(|V_2|, q)$. By our assumptions, for sufficiently large n,

$$q \gg \max\left\{\frac{\ln|V_1|}{|V_1|}, \frac{\ln|V_2|}{|V_2|}\right\},$$

   so both G_1 and G_2 are connected with probability approaching 1. In this case, the given procedure has a unique, *correct* choice, and thus it will identify the correct partition with probability tending to 1 as $n \to \infty$.