Randomness and Computation 2017/18
Coursework 2 (summative)

Issue date: Sunday, 4th March, 2018

The deadline for this coursework is 4pm on Monday, 19th March 2018 (Monday of week 9). Please submit your solutions either *electronically via* `submit` *or by hand to the ITO in Appleton Tower.* **Remember that the School's policy on late coursework means that late coursework incurs a penalty unless you have good justification for the lateness (as per the "good reasons" definitions on our webpages). If you believe you do satisfy the "good reasons" conditions, you should request an extension (well in advance) through the ITO.**

This coursework should be your own individual work. *You may discuss understanding of the questions with your classmates, but may not share solutions, or give strong hints. If you use any resources apart from the course slides/notes, or the book, you* must *cite these in detail (on a per-question basis.)*

This coursework is worth 20% of your final course average.

1. We consider the number of *acyclic orientations* in (undirected) Erdős-Rényi graphs.

   Definition 1 *For any directed graph* $\vec{G} = (V, \vec{E})$*, we say that* $\vec{G}$ *is* acyclic *if there is no directed cycle in* $\vec{G}$*. Such a graph is sometimes called a* directed acyclic graph (DAG)*.*

   Definition 2 *For any given (simple) undirected graph* $G = (V, E)$*, an* orientation $\vec{G}$ *of* G *is any directed graph* $\vec{G} = (V, \vec{E})$ *such that* $|\vec{E}| = |E|$ *and such that for every* $(u, v) \in E$*, exactly* one *of the arcs* $(u \to v)$ *and* $(v \to u)$ *belongs to* $\vec{E}$*.*

   Definition 3 *For any simple* undirected *graph* $G = (V, E)$*, the set of* acyclic orientations (AOs) *of* G*, denoted* $AO(G)$*, is the set of all orientations* $\vec{G}$ *of* G *which are acyclic. Such a graph is sometimes called a* directed acyclic graph (DAG)*.*

   The problem of *counting the number of acyclic orientations* of a given undirected graph is a well-known $\sharp$P-complete ("hard to count") problem. We will consider the problem when the input graph is drawn from the *random model* $\mathcal{G}_{n,p}$ and show how to evaluate the *expected* number of AOs in polynomial-time. Unlike some simple structures that we will count in $G_{n,p}$ in our lectures (eg, clique on 4 vertices), we will not prove an exact value for $\mathrm{E}_{n,p}[|AO(G)|]$; instead we will derive an algorithm to compute the expectation, given n, p.

   Definition 4 *The Erdős-Rényi model* $\mathcal{G}_{n,p}$ *of random graphs is parametrized by the number of vertices* n*, and an edge addition probability* $p \in [0, 1]$*.*

   We generate a undirected simple graph $G = (V, E)$ *from* $\mathcal{G}_{n,p}$ *by setting* $V = \{1, \ldots, n\}$*. To construct the edge set* E*, we consider each* (i, j) *pair* $1 \le i < n$*,* $i < j \le n$ *independently, and we add the undirected edge* (i, j) *to* E *with probability* p *(omitting it with probability* $1 - p$*).*

The Erdős-Rényi process always creates a simple graph without loops or parallel edges. The particular number of edges added will vary depending on the results of the independent trials for the edges. The *expected* number of edges, written $\mathbb{E}_{n,p}[|E|]$ has the value $p \cdot \binom{n}{2}$. Also, for a particular "number of edges" m which might be generated, all simple graphs with m edges have the same probability (which is $p^m(1-p)^{\binom{n}{2}-m}$) in the Erdős-Rényi model for n, p.

We now present a recurrence which was proved independently by two famous combinatoricists in the early 1970s. We need one more definition first:

Definition 5 *Let $n, m \in \mathbb{N}$. The number of different* directed acyclic graphs (DAGs) *on n vertices with m arcs, is denoted by $A_{n,m}$.*

We never compute the $A_{n,m}$ values; we just need to work with the definition.

Theorem 6 (Robinson, Stanley) *Suppose we define the quantity $A_n(x)$ by*

$$A_n(x) = \sum_{m=0}^{\binom{n}{2}} A_{n,m} x^m.$$

(the co-efficients of $A_n(x)$ are the counts of DAGs with m arcs for various m). Then

$$A_n(x) \quad = \quad \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} (1+x)^{i(n-i)} A_{n-i}(x).$$

We will *not* be proving the correctness of this recurrence, as it is difficult. We will be *applying* it to compute the *expected number of AOs* of a random graph G drawn from $G_{n,p}$.

The expected number of AOs for a random graph from $\mathcal{G}_{n,p}$ is defined as

$$\mathbb{E}_{n,p}[|AO(G)|] = \sum_{G=(V,E),|V|=n} \mathbb{P}r_{n,p}[G] \cdot |AO(G)|,$$

where $\mathbb{P}r_{n,p}[G]$ is the probability that G is generated by $\mathcal{G}_{n,p}$.

(a) Prove that the Robinson-Stanley polynomial $A_n(x)$, when evaluated at $x = \frac{p}{1-p}$ for any $p \in (0, 1)$, satisfies the following equality: *[15 marks]*

$$A_n\left(\frac{p}{1-p}\right) = (1-p)^{-\binom{n}{2}} \times \mathbb{E}_{n,p}[|AO(G)|].$$

You will not need to use the result of Theorem 6 for this part, just the definition of $A_n(x)$. You will not need to use *induction*, or *proof by contradiction*, just manipulation and simplification of equations. The proof will only be about 5-6 lines long if done right.

(b) Design an $O(n^2)$-time (or at worst, an $O(n^3)$-time) algorithm) to evaluate $\mathbb{E}_{n,p}[|AO(G)|]$ exactly for given values n, p. Justify the running time in detail. *[15 marks]*

The two main facts you will need for this are the result of part (a) (though not necessarily the proof of it), and the recurrence of Thm 6.

2. In Question 4 of Coursework 1 this year, we worked to develop a variation of the "Max Cut" algorithm from Lecture 4 which gives a slightly better expectation for the size of the random cut of a graph $G = (V, E)$, by ensuring that the random cut splits the vertex set in half (or almost half, if $|V|$ is odd). Let $n = |V|$ and $m = |E|$ as usual. The new algorithm fixes the size of the "left side" to be $\lfloor \frac{n}{2} \rfloor$, and then samples vertices *without replacement* from V until the "left side" is full, taking the leftover vertices for the right-hand side. In our work on coursework 1 we were able to show that for this algorithm, the expected number of edges crossing the cut will be the slightly larger value $\frac{n}{2n-1}m$ rather than the $\frac{m}{2}$ of our original algorithm from Lecture 4.

   In Lecture 10 we have already seen an example of how we can "derandomize" a process to get a deterministic algorithm/method which is guaranteed to return a solution which is at least as good as the expectation from the randomized algorithm (the example was for 2 edge-colourings of K_n which avoid monochromatic K_4 subgraphs). There is also an example of how to derandomize the original Max Cut algorithm in Section 6.2 of our course text by Mitzenmacher and Upfal.

   Show how to derandomize our improved algorithm described above. Your algorithm should be low polynomial-time (something like $O(n^2)$ or $O(n^3)$). Justify the fact that your method will *definitely* return a cut with at least $\frac{n}{2n-1}m$ edges, using appropriate reference to conditional expectations. *[15 marks]*

   Note that you don't need to have done Question 4 to be able to continue with this part. The information in this question is sufficient.

3. Suppose we have access to some Bernoulli random variable Y with unknown parameter $p \in (0, 1)$ ($p = \Pr[Y = 1]$), and we want to solve two problems using repeated sampling from Y, given an natural number N as input:

   (a) To compute a value $\widehat{p} \in (0, 1)$ such that $\Pr\left[|\widehat{p} - p| \leq \frac{1}{N}\right]$ is at least $1 - \frac{1}{N}$; *[10 marks]*

   (b) To determine, with probability at least $1 - \frac{1}{N}$, whether $p \leq \frac{1}{4}$ or not. *[10 marks]*

   For each of these two problems, determine whether the result can be achieved by drawing a polynomial-number of samples from Y. If the result is possible, compute the minimum number of samples needed to guarantee the result, subject to the use of the Chernoff bounds of Corollary 4.6. If you think the result cannot be achieved with polynomially-many samples, give your reasons.

4. Again consider the Erdős-Rényi model $G_{n,p}$ model of random graphs (where every edge is included independently and identically with probability p).

   We will be interested in the presence (or absence) of a *4-cycle* in the random graph - a 4-cycle being a set of four vertices which can be arranged so that all 4 outer induced edges belong to the generated graph G, but neither of the two "crossing edges" belong to G. We will consider the set of indicator variables $\{X_f : f \subseteq [n], |f \cap [n]| = 4\}$, where X_f will be 1 for $f = \{u_f, v_f, w_f, z_f\}$ iff these vertices can be ordered so that the 4 "outer edges" lie in the generated graph $G \leftarrow G_{n,p}$, but the two "crossing edges" are absent from G.

   We will be interested in the total number of 4-cycles $X = \sum_{f\subseteq[n],|f\cap[n]|=4} X_f$, and the expectation and variance of X.

   (a) Derive an exact expression for the expected number of 4-cycles $\mathrm{E}[X]$ in the random graph. This will depend on n and p. *[5 marks]*

   (b) Using the $\mathrm{E}[X]$ value from (b), show that for any $p = p(n)$ that satisfies $pn \to 0$ as $n \to \infty$, that $\Pr[X > 0] \to 0$. *[5 marks]*

   (c) Derive a close estimate of the second moment $\mathrm{E}[X^2]$, taking into account the different cases for pairs of 4-cycles to overlap or not-overlap. Hence derive a bound on the variance $\mathrm{Var}[X]$. *[10 marks]*

   (d) Using your result from (c), use the special case of Chebyshev's Inequality with $a = \mathrm{E}[X]$ to derive a sufficient condition on $p = p(n)$ (with respect to n) to ensure $\Pr[X > 0] \to 1$ (equivalently, that $\Pr[X = 0] \to 0$). *[5 marks]*

   This question requires most of the content of Lectures 10-12 on "the probabilistic method".

5. We are given an undirected graph $G = (V, E)$ where each $v \in V$ is associated with a set of $8r$ colours $S(v)$, for some $r \geq 1$.

   The $S(v)$ sets may overlap or in some cases be identical, or anything in between; however we have the guarantee that for every $v \in V$ and every $k \in S(v)$, there are at most r neighbours $u \in \mathsf{Nbd}(v)$ that also have $k \in S(u)$.

   Prove that there is a proper (vertex) colouring which assigns a colour from $S(v)$ to each $v \in V$ such that for every $e = (u, v) \in E$, u and v get different colours. *[10 marks]*

   (in solving this, it may be helpful to define a collection of events $\{A_{u,v,k}\}$, with $A_{u,v,k}$ representing the event that both u and v get the colour k)

   This question requires most of the content of Lectures 10-12 on "the probabilistic method".

Please submit your work in advance of 4pm, Monday 19th March *either* in person to the ITO (in Appleton Tower), *or* electronically from DICE using the following `submit` command (for your file named `coursework2.pdf`):

```
submit rc cw2 coursework2.pdf
```

Mary Cryan, 4th March 2018