Randomness and Computation 2018/19
Solutions and marking scheme for Coursework 2

Solutions release date: Monday, 7th April, 2019

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

   (a) We now 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]*

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

We are given the following definition for the polynomial $A_n(x)$:

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

where $A_{n,m}$ is the *number of directed acyclic graphs (DAGs) on* n *nodes and having* m *arcs.*

We use the definition of $A_n(\cdot)$ applied to $x = \frac{p}{1-p}$ into $A_n(x)$ and manipulating the expression:

$$A_n\left(\frac{p}{1-p}\right) = \sum_{m=0}^{\binom{n}{2}} A_{n,m}\left(\frac{p}{1-p}\right)^m. \quad (1)$$

We can re-write $(\frac{p}{1-p})^m$ as $(1-p)^{-\binom{n}{2}} p^m (1-p)^{\binom{n}{2}-m}$ for any value $m = 0, \ldots, \binom{n}{2}$, to give the following:

$$A_n\left(\frac{p}{1-p}\right) = \sum_{m=0}^{\binom{n}{2}} A_{n,m}(1-p)^{-\binom{n}{2}} p^m (1-p)^{\binom{n}{2}-m} \quad (2)$$

$$= (1-p)^{-\binom{n}{2}} \sum_{m=0}^{\binom{n}{2}} A_{n,m} p^m (1-p)^{\binom{n}{2}-m}. \quad (3)$$

At this point, we observe that $p^m(1-p)^{\binom{n}{2}-m}$ is the probability that the particular undirected graph G with m edges is generated in $\mathcal{G}_{n,p}$.

Now consider $A_{n,m}$, the number of Directed Acyclic Graphs with n nodes, m arcs. Observe:

(i) For every DAG $\vec{G}$ (with n nodes, m arcs), there is exactly *one* undirected graph G (which will have n nodes, m edges) such that $\vec{G} \in AO(G)$.

(ii) The set of all Directed Acyclic Graphs (DAGs) with n nodes and m arcs is equal to

$$\bigcup_{G:|V|=n,|E|=m} AO(G).$$

Hence we can replace $A_{n,m}$ with $|\bigcup_{G:|V|=n,|E|=m} AO(G)| = \sum_{G:|V|=n,|E|=m} |AO(G)|$. Hence

$$A_n\left(\frac{p}{1-p}\right) = (1-p)^{-\binom{n}{2}} \sum_{m=0}^{\binom{n}{2}} \left(\sum_{G:|V|=n,|E|=m} |AO(G)| \right) p^m (1-p)^{\binom{n}{2}-m} \quad (4)$$

$$= (1-p)^{-\binom{n}{2}} \sum_{m=0}^{\binom{n}{2}} \left(\sum_{G:|V|=n,|E|=m} |AO(G)| p^m (1-p)^{\binom{n}{2}-m} \right), \quad (5)$$

$$= (1-p)^{-\binom{n}{2}} \sum_{m=0}^{\binom{n}{2}} \left(\sum_{G:|V|=n,|E|=m} \mathbb{P}r_{n,p}[G] \cdot |AO(G)| \right), \quad (6)$$

where this recent step follows from the fact that any particular graph G with n nodes and m edges has probability $p^m(1-p)^{\binom{n}{2}-m}$ of being generated by $\mathcal{G}_{n,p}$. Now m does not appear inside the expression $\mathbb{P}r_{n,p}[G] \cdot |AO(G)|$. Hence we can remove the partition wrt m to get

$$A_n\left(\frac{p}{1-p}\right) = (1-p)^{-\binom{n}{2}} \sum_{G:|V|=n} \mathbb{P}r_{n,p}[G] \cdot |AO(G)|, \quad (7)$$

as required.

marking: The 15 are shared as follows:

- 1 mark for writing $A_n(x)$ out at $x = \frac{p}{1-p}$ (eq (1)).
- 3 marks for re-writing $(\frac{p}{1-p})^m$ as $(1-p)^{-\binom{n}{2}} p^m (1-p)^{\binom{n}{2}-m}$ and moving the $(1-p)^{-\binom{n}{2}}$ term to the outside of the sum (eq (3)).
- 2 marks for observing in words that the probability of generating G in $\mathcal{G}_{n,p}$ is $p^m(1-p)^{\binom{n}{2}-m}$ for any particular G with m edges.
- 3 marks for the observation (i). I want this spelled out in sentences, with references to the definitions of $A_{m,n}$ and $AO(G)$.
- 3 marks for writing $A_{n,m}$ as the sum of $|AO(G)|$ over graphs G with n nodes and m edges, ie applying (ii) (to get eq (4)/(5)).
- 3 marks for noticing that $p^m(1-p)^{\binom{n}{2}-m}|AO(G)|$ is the contribution of graph G to the expectation, and working to $\sum_{G:|V|=n} \mathbb{P}r_{n,p}[G] \cdot |AO(G)|$ (eqs (6) and (7)).

(b) We now design an $O(n^2)$-time algorithm) to evaluate $\mathbb{E}_{n,p}[|AO(G)|]$ exactly for given values n, p, and justify its running time. *[10 marks]*

By the result of (a) we can take the approach of computing $A_n(\frac{p}{1-p})$, and the multiplying that value by $(1-p)^{\binom{n}{2}}$.

In setting up the computation of $A_n(\frac{p}{1-p})$, observe that the recurrence only involves calls to smaller n *with the same* $x = \frac{p}{1-p}$. Hence we only need a one-dimensional "dynamic programming table" (where the index will specify the "n" index). We will use $A[\widehat{n}]$ to store the value of $A_{\widehat{n}}(\frac{p}{1-p})$ for $\widehat{n} = 0, \ldots, n$. We will evaluate $A_{\widehat{n}}(\frac{p}{1-p})$ in order of increasing $\widehat{n}$, applying the recurrence for each evaluation. We will apply this recurrence *more or less directly*, except that during this application, we will find $(-1)^{i+1}$, $\binom{\widehat{n}}{i}$ and $(1+\frac{p}{1-p})^{i(\widehat{n}-i)}$ using a "better-than-naive". Specifically, we will compute $(-1)^{i+1}$ using a single multiplication operation multiplying the previous value by (-1); similarly we will compute $\binom{\widehat{n}}{i}$ using one multiplication and a division, exploiting the relationship $\binom{\widehat{n}}{i} = \binom{\widehat{n}}{i-1}\frac{\widehat{n}-i+1}{i}$. Finally, we will *look-up* the $(1+\frac{p}{1-p})^{i(\widehat{n}-i)}$ values from a linear array which allows direct index to any power (below $\lceil\frac{n^2}{4}\rceil$). This array, named $\mathit{allpows}$, is computed in $\Theta(n^2)$ at the beginning of the Algorithm.

Here is the pseudocode:

Algorithm EXP-NP(n, p)

double $A[n+1]$
double $\mathit{allpows}[\lceil\frac{n^2}{4}\rceil + 1]$
int binom, i, sign
$\mathit{allpows}[0] \leftarrow 1$
for $i \leftarrow 1$ **to** $\lceil\frac{n^2}{4}\rceil$
    $\mathit{allpows}[i] \leftarrow \mathit{allpows}[i-1] \cdot \frac{1}{1-p}$
$A[0] \leftarrow 1, A[1] \leftarrow 1$
for $\mathit{nhat} \leftarrow 2$ **to** n
    $A[\mathit{nhat}] \leftarrow 0$
    $\mathit{binom} \leftarrow 1$
    $\mathit{sign} \leftarrow -1$
    for $i \leftarrow 1$ **to** nhat
        $\mathit{sign} \leftarrow \mathit{sign} \cdot (-1)$
        $\mathit{binom} \leftarrow \mathit{binom} \cdot (\mathit{nhat} - i + 1)/i$
        $A[\mathit{nhat}] \leftarrow A[\mathit{nhat}] + \mathit{sign} \cdot \mathit{binom} \cdot \mathit{allpows}[i \cdot (\mathit{nhat} - i)] \cdot A[\mathit{nhat} - i]$
return $(1-p)^{\binom{n}{2}} \cdot A[n]$

For $\Theta(n^2)$ running time:

- Lines 1, 2, 3 are just declarations, $\Theta(1)$ time (or $\Theta(n^2)$ if they assume explicit visiting of all array cells during declaration).
- Line 4 is $\Theta(1)$.
- The loop in lines 5-6 runs $\lceil \frac{n^2}{4} \rceil$ times, doing a single multiplication at each step ... $\Rightarrow \Theta(n^2)$ time for the entire loop.
- Line 7 $\Theta(1)$ time
- Loop in lines 8-15 runs $n - 1$ times, for each iteration for "nhat":
  - Lines 9, 10, 11 take $\Theta(1)$ time.
  - Then the *inner loop* runs for i = 1, ..., nhat, at each step doing:
    (in line 13) 1 mult,
    (in line 14) 1 mult, 1 addition, 1 subtract and 1 division, and
    (in line 15) 2 direct array lookups, 3 mults and 1 addition.
    So $\Theta(1)$ work for every $i = 1, \ldots, nhat \ldots$
    $\Rightarrow$ *Inner loop* does $\Theta(nhat)$ work overall.
- So the *Outer loop* in lines 8-15 takes $\sum_{nhat=1}^{n} \Theta(nhat)$ time overall, which is $\Theta(n^2)$.
- Line 16 certainly takes (at most) $\Theta(n^2)$ time.

So overall $\Theta(n^2)$.

marking: 7 marks are going for Algorithm details, plus 3 for analysis of running-time.

- Algorithm:
  - Up to 7 marks for setting out any correct application of the Stanley-Robinson recurrence for $p/(1-p)$ which uses (linear) DP to calculate $A_n(p/(1-p))$ *and then* multiplies by $(1-p)^{\binom{n}{2}}$.
  - Penalize small errors or omissions by 1 or 2 marks.
- Justification of running time (3 marks):
  - If they can only show $\Theta(n^3)$, work from 2 marks.
  - Give 3 marks for a decent $\Theta(n^2)$ justification.
    They should in particular have the "allpows" array approach - 1 of the 3 marks go for spelling out how to deal with all the powers inside this array, and using lookup later.
    2 of the 3 for the various details of getting $\Theta(n^2)$. They don't have to have given all the detail I did (but should do some). I think that quite a few of them built the set of binomial coefficient values using "Pascal's triangle" (but that approach also has the $\Theta(n^2)$ time so it's fine).

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

   (a) We have access to a Bernoulli random variable Y with unknown parameter $p \in (0, 1)$ *[10 marks]*

($p = \Pr[Y = 1]$), and we want to compute a value $\widehat{p} \in (0, 1)$ such that $\Pr\left[|\widehat{p} - p| \le \frac{1}{N}\right]$ is at least $1 - \frac{1}{N}$.

The natural approximation is to draw many samples from Y (say $Y_1, \ldots, Y_n$, sum them all as $\mathcal{Y} = \sum_{i=1}^{n} Y_i$) and take their average, $\mathcal{Y}/n$, which will have expectation exactly p. Our aim is have n sufficiently large to make $\Pr\left[|\widehat{p} - p| \le \frac{1}{N}\right]$ large enough. We recast this probability in terms of $\mathcal{Y}$, to note that this probability is exactly $\Pr\left[|\mathcal{Y} - pn| \le \frac{n}{N}\right]$ (recall $E[\mathcal{Y}] = np$). We want to make this probability at least $\frac{1}{N}$, equivalent to making $\Pr\left[|\mathcal{Y} - pn| > \frac{n}{N}\right]$ less than $\frac{1}{N}$. Note that at this point we don't know how large n will need to be to achieve this.

Let us now view this probability in relation to Chernoff bounds, using the bound of Corollary 4.6 which implies that for any $\delta, 0 < \delta < 1$,

$$\Pr[|\mathcal{Y} - np| \ge \delta np] \;\le\; 2e^{-np\delta^2/3}.$$

Note this is bounding $\Pr[|\mathcal{Y} - np| \ge \delta np]$, but then it *certainly* also bounds $\Pr[|\mathcal{Y} - np| > \delta np]$ which is the form of our probability of interest.

Now to bound deviation $\mathcal{Y}$ to within $\frac{n}{N}$, we want to have $\delta np = \frac{n}{N}$, which happens if $\delta = \frac{1}{Np}$. By Corollary 4.6 this would happen with probability at most $2e^{-np/(3N^2p^2)}$, which is $2e^{-n/(3N^2p)}$. We want to have $2e^{-n/(3N^2p)} < \frac{1}{N}$, which will happen if and only if $2N < e^{n/(3N^2p}$, ie if $\ln(2N) < \frac{n}{3N^2p}$, ie if and only if $n > 3N^2p\ln(2N)$. We don't know what the value of p is (of course), but it's at most 1, so we can take any $n > 3N^2\ln(2N)$ as our number of samples, which is definitely polynomial in the given N.

marking: 3 marks are going for the reformulation of our target result to $\Pr[\mathcal{Y} - pn| > \frac{n}{N}$, and I'm expecting at least some discussion to get all those 3 marks. Next 3 marks going for the comparison with Chernoff (corollary 4.6) and the calculation of what δ has to be. Then 3 more marks for working with the $2e^{-\cdots}$ compared to $\frac{n}{N}$ and working out a (tight) bound on n which will suffice to get our result. Final mark is going for either the discussion of the $\le, <$ issue *or* everything being really brilliant.

extra note: In the use of inequality 2. of the Chernoff bounds I am (with stating it) implicitly assuming that $\frac{1}{N} < p$ (to have the $\delta < 1$ condition. I forgot to write down the details for when $N \ge p^{-1}$, and I didn't penalise anyone who forgot this step. However, first note that the $\frac{1}{N} < p$ is the more interesting case (where we look for a relatively tight bound). For the $\frac{1}{N} \ge p$ case we first notice that if $\frac{1}{N} \ge 5p$, then we can use part 3. of Theorem 4.4 to show that the probability of $\mathcal{Y} - np \ge \frac{1}{N}$ is at most 2^{-5pn}, which is $\le 2^{-5n/N}$, and this bound will be less than $\frac{1}{N}$ when $n \ge \frac{1}{5}N\ln(N)$. Of course this leaves the case when we have N with $p < \frac{1}{N} < 5p$; well in that case, we will just take enough samples to achieve a bound within $\lfloor p^{-1} \rfloor$ (hence better than $\frac{1}{N}$), and this will give rise to a couple of factors 5 when we convert the bound in relation to the original given N.

(b) We have access to a Bernoulli random variable Y with unknown parameter $p \in (0, 1)$ ($p = \Pr[Y = 1]$), and we want to determine, with probability at least $1 - \frac{1}{N}$, whether $p \le \frac{1}{4}$ or not. *[10 marks]*

For this question, consider the situation where we have $p = 1/4$ exactly. Then if we define $\mathcal{Y} = \sum_{i=1}^{n} Y_i$ for large n as in part (i), then $E[\mathcal{Y}/n] = 1/4$. However, the observed value of $\frac{\mathcal{Y}}{n}$ will vary on either side of $\frac{1}{4}$.

Suppose there was a polynomial $p(N)$ which would answer whether $p \leq \frac{1}{4}$ using $p(N)$ samples, and imagine we had both two Bernoulli's Y and Y', Y with $p = 1/4$ exactly, and Y' with $p = 1/4 + 1/(4p(N)^2)$. Then the expected value $E[\mathcal{Y}]$ will be $\frac{p(N)}{4}$ and the expected value $E[\mathcal{Y}']$ will be $\frac{p(N)}{4} + \frac{1}{4p(N)}$. The difference in these values is far smaller than 1, and hence, the probability we would observe a higher value for $\mathcal{Y}'$ than $\mathcal{Y}$ is very small, and *provably* less than $\frac{1}{4}$. Hence the probability we could distinguish between those two situations ($p = 1/4$ and $p = 1/4 + 1/(4p(N)^2)$) certainly cannot reach $1 - \frac{1}{N}$ for general N. Note we were able to design the Y' variable after knowing the chosen polynomial $p(N)$, hence increasing $p(N)$ to a different polynomial will not help.

marking: If they just mention p might be close to $1/4$, and that we would not be able to decide whether it was greater or less using reasonable samples, then 3-4 marks.

In order to get all 10 marks they should be giving an argument like mine, with a specific competitor Y' that is very close, and describing the effect it has on $E[\mathcal{Y}], E[\mathcal{Y}']$. They might not get quite as far as I did (the particular choice of the $\frac{1}{4p(N)^2}$ means we can prove a rigorous bound of $\frac{1}{4}$ on how often we'd see a difference), so give 6-7 marks for a weaker argument.

3. We are analysing the likelihood of *4-cycles* in $G_{n,p}$ - 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.

   (a) We first derive an exact expression for the expected number of 4-cycles $E[X]$ in the random graph. *[5 marks]*

   Recall that we consider every 4-set of vertices $\{u, v, w, z\}$ and analyse whether this is a 4-cycle or not. This will be the case if and only if we have u adjacent to two of the other vertices, but *not* the third, and that third one adjacent to the "two", but *not* to u. Not that because of the regularities, we could have chosen any of the 4 vertices to play the role of u.

   The indicator variable for success of $\{u_f, v_f, w_f, z_f\}$ is X_f and has expectation $E[X_f] = 3[(1-p)p^2]^2 = 3(1-p)^2p^4$, with 3 being the choice of the "third" non-u vertex, with $p^2(1-p)$ being the probability that u has the correct edge presence/absence with "the two" and the "third", and the second $p^2(1-p)$ representing the same for the "third" vertex.

   There are $\binom{n}{4}$ ways to choose a 4-set of vertices from the graph (and hence $\binom{n}{4}$ 4-sets), then using linearity of expectation with our analysis of a single X_f, we have $E[X] = 3\binom{n}{4}p^4(1-p)^2$.

marking: They should do a careful (correct) analysis for $E[X_f]$ for a particular f, with justification, up to 3 marks for this. Other 2 marks for for mentioning linearity of exp and using it to get the overall value $E[X]$.

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

First note that since X is a non-negative random variable on the integers, that $\Pr[X > 0] = \Pr[X \geq 1]$ and by Markov's Inequality, $\Pr[X \geq 1] \leq \frac{E[X]}{1} = E[X]$.

Now

$$E[X] = 3\frac{n(n-1)(n-2)(n-1)}{24}p^4(1-p)^2 = \frac{n(n-1)(n-1)(n-3)}{8}p^4(1-p)^2.$$

If we have $pn \to 0$ as $n \to \infty$, then $(1-p) \to 1$. Then

$$E[X] = \frac{(1-p)^2}{8}(np)((n-1)p)((n-2)p)((n-3)p) \to \frac{1}{8} \cdot 0 \cdot 0 \cdot 0 \cdot 0 \;=\; 0$$

as $n \to \infty$. Hence $\Pr[X > 0] \to 0$ as $n \to \infty$ also.

marking: 2 marks for using/explaining Markov's inequality, then the remaining three for a few intermediate steps showing how it can be applied to get the $\to 0$.

(c) We now derive a close estimate of the second moment, $E[X^2]$. *[10 marks]*

Note

$$E[X^2] \;=\; E[\sum_f \sum_g X_f X_g] \;=\sum_f \sum_g E[X_f X_g],$$

where f, g are each taken over the 4-sets of V, and where we use linearity of expectation in the second step. We will evaluate $E[X_f X_g]$ for arbitrary pairs of 4-sets f and g, breaking our analysis into cases.

case (i): $|f \cap g| \leq 1$. In this case, none of the edges to be considered for X_f overlap with edges influencing X_g. Hence those r.v s are independent, and $E[X_f X_g] = E[X_f]E[X_g]$.

case (ii): $|f \cap g| = 2$. In this case, *one of the potential edges* (of six) influencing X_f overlaps with one of the potential edges influencing g, with the other (5 apiece) edges being disjoint.

Let u, v be the two vertices that lie in both f and g, and let the disjoint vertices be w_f, z_f for f and w_g, z_g for g.

It might be the case that the overlapping potential edge $\{u, v\}$ gets added to G. This happens with probability p. Then conditional on $\{u, v\} \in G$, we consider the options to achieve X_f by evaluating what happens with w_f, z_f and their shared edges to u, v. If $\{u, v\}$ is already in G, then to have X_f equal to 1, we must *either* have $\{u, w_f\} \in G, \{u, z_f\} \notin G, \{v, w_f\} \notin G, \{v, z_f\} \in G, \{w_f, z_f\} \in G$ *or* alternatively have $\{u, w_f\} \in G, \{u, z_f\} \notin G, \{v, w_f\} \notin G, \{v, z_f\} \in G, \{w_f, z_f\} \in G$. Each of these two conditional events has probability $p^3(1-p)^2$, hence $\Pr[X_f = 1 \mid \{u, v\} \in G] = 2p^3(1-p)^2$. We have the exact same (independent) conditional probability for $\Pr[X_g = 1 \mid \{u, v\} \in G]$; hence $\Pr[X_f X_g = 1 \mid \{u, v\} \in G] = [2p^3(1-p)^2]^2$.

Alternatively, it might be the case that the overlapping potential edge does not succeed in being added to G (with probability $1-p$). Conditional on $\{u,v\} \notin G$, we analyse the probability of X_f by noticing that u and v must be on opposite "sides of the cycle, and so must z_f and w_f (ie, we must have $\{w_f, z_f\} \notin G$). Also we require all of the edges $\{u,z_f\},\{u,w_f\},\{v,z_f\},\{v,w_f\}$ to be added to G. Hence, the probability $\Pr[X_f = 1 \mid \{u,v\} \notin G] = (1-p)p^4$. We have the same (independent) condition probability for $\Pr[X_g = 1 \mid \{u,v\} \notin G]$; hence $\Pr[X_f X_g = 1 \mid \{u,v\} \notin G] = [(1-p)p^4]^2$.

Putting everything together, $E[X_f X_g] = p[2p^3(1-p)^2]^2 + (1-p)[(1-p)p^4]^2$, which is $4p^7(1-p)^4 + p^8(1-p)^3 = p^7(1-p)^3[4(1-p)+p]$.

case (iii): $|f \cap g| = 3$. In this case, *three of the potential edges* (of six) influencing X_f overlaps with three of the potential edges influencing g, with the other (3 apiece) edges being disjoint.

Let u, v, w be the two vertices that lie in both f and g, and let the disjoint vertices be z_f for f and z_g for g.

Examining what happens on u,v,w, note that if we view an induced cycle on $\{u,v,w,z_f\}$, then removing z_f temporarily will induce a 2-edge path on the other three vertices, *without* the closing edge of a triangle. In order for a configuration on $\{u,v,w\}$ to be extending to an induced 4-cycle on the 4 vertices, $\{u,v,w\}$ must have 2 of the possible edges added, the 3rd omitted. The probability of this in $G_{n,p}$ is exactly $3p^2(1-p)$. Then, when we consider the completion to a 4-cycle including z_f, there is only one configuration that will achieve this, which is adding edges from z_f to the two endpoints of the 2-edge path, and omitting the edge form z_f to the middle vertex. Hence the conditional probability of completing the 2-edge path to an induced 4-cycle is $p^2(1-p)$. The same is true (independently) for the conditional probability to achieve an induced 4-cycle with z_g. Hence $\Pr[X_f X_g] = 3p^2(1-p)[p^2(1-p)]^2 \ = \ 3p^6(1-p)^3$.

case (iv): Of course for $|f \cap g| = 4$, $E[X_f X_g] = E[X_f]$.

Now putting everything together, we note that there are $\binom{n}{4}\binom{n-4}{4} + n \cdot \binom{n-1}{3}\binom{n-4}{3}$ pairs f, g satisfying case (i), there are $\binom{n}{2}\binom{n-2}{2}\binom{n-4}{2}$ pairs satisfying case (ii), there are $\binom{n}{3}(n-3)(n-4)$ pairs satisfying case (iii), and $\binom{n}{4}$ pairs satisfying case (iv). Note that the contribution to $E[X^2]$ from all these $\binom{n}{4}$ matching pairs is of course $E[X]$.

Overall,

$$
\begin{aligned}
E[X^2] \quad &= \quad \left(\binom{n}{4}\binom{n-4}{4} + n\binom{n-1}{3}\binom{n-4}{3}\right)[9p^8(1-p)^4] + \\
&\binom{n}{2}\binom{n-2}{2}\binom{n-4}{2}[p^7(1-p)^3(4(1-p)+p)] + \binom{n}{3}(n-3)(n-4)3p^6(1-p)^3 + E[X]
\end{aligned}
$$

Let us now square $E[X]$, this is $E[X]^2 = \binom{n}{4}^2 9p^8(1-p)^4$. Recall that $\mathrm{Var}[X] = E[X^2] - E[X]^2$, and let us subtract $E[X]^2$ from the initial term of $E[X^2]$. The difference in the binomial terms for $9p^8(1-p)^4$ is

$$
\begin{aligned}
& \binom{n}{4}\binom{n-4}{4} + n\binom{n-1}{3}\binom{n-4}{3} - \binom{n}{4}^2 \\
= & \binom{n}{4}\left(\binom{n-4}{4} - \binom{n}{4}\right) + n\binom{n-1}{3}\binom{n-4}{3} \\
= & \binom{n}{4}\left(\binom{n-4}{4} - \binom{n}{4}\right) + 4\binom{n}{4}\binom{n-4}{3} \\
= & \binom{n}{4}\left(\binom{n-4}{4} + 4\binom{n-4}{3} - \binom{n}{4}\right) \\
= & \binom{n}{4}\left(\binom{n-3}{4} - \binom{n}{4}\right) \\
< & \ 0.
\end{aligned}
$$

So $\mathrm{Var}[X] < \binom{n}{2}\binom{n-2}{2}\binom{n-4}{2}[p^7(1-p)^3(4(1-p)+p)] + \binom{n}{3}(n-3)(n-4)3p^6(1-p)^3 + \mathrm{E}[X]$

marking: 2 marks explaining how we need to evaluate $\mathrm{E}[X_f X_g]$ for general 4-sets f, g, and that this expectation will vary according to independence of the edges relevant to X_f, X_g. Then 1 mark for case (i), 2 for case (ii), 2 for case (iii), 1 for case (iv). The final 2 marks for wrapping it all up to get a sensible overall bound on $\mathrm{E}[X^2], \mathrm{Var}[X]$ (they don't need to have it with $O(\cdot)$).

This is a hard question so give some partial marks for imperfect solutions.

(d) We now note that $\Pr[X = 0] = \Pr[X - \mathrm{E}[X] = -\mathrm{E}[X]] < \Pr[|X - \mathrm{E}[X]| \geq \mathrm{E}[X]]$. We can apply Chebyshev's Inequality, which states $\Pr[|X - \mathrm{E}[X]| \geq \mathrm{E}[X]] \leq \frac{\mathrm{Var}[X]}{\mathrm{E}[X]^2}$. Hence *[5 marks]*

$$
\begin{aligned}
\Pr[X = 0] \quad & < \quad \frac{\binom{n}{2}\binom{n-2}{2}\binom{n-4}{2}[p^7(1-p)^3(4(1-p)+p)] + \binom{n}{3}(n-3)(n-4)3p^6(1-p)^3}{\binom{n}{4}^2 9p^8(1-p)^4} + \frac{1}{3\binom{n}{4}p^4(1-p)^2} \\
& = \quad \frac{O(n^6p^7(1-p)^3) + O(n^5p^6(1-p)^3))}{\binom{n}{4}^2 9p^8(1-p)^4} + \frac{1}{3\binom{n}{4}p^4(1-p)^2} \\
& = \quad O\left(\frac{1}{n^2p(1-p)}\right) + O\left(\frac{1}{n^3p^2(1-p)}\right) + O\left(\frac{1}{n^4p^4(1-p)^2}\right)
\end{aligned}
$$

We are asked to give conditions on $p = p(n)$ to ensure $\Pr[X > 0] \to \infty$. That's equivalent to looking for conditions to ensure $\Pr[X = 0] \to 0$, and this means we need to make sure each denominator needs to tend to infinity as $n \to \infty$. This will require p and $(1-p)$ to be bounded away from 0, specifically we want to set p so that each of $n^2p(1-p), n^3p^2(1-p), n^4p^4(1-p)^2$ tends to ∞. For p closer to 0 than 1, this means we need $(np) \cdot n$, $(np)^2 n$ and $(np)^4$ all to tend to infinity, hence we need $np \to \infty$. For p close to 1, then we need to ensure $(1-p)n^2, (1-p)n^3$ and $(1-p)^2 n^4$ to tend to infinity, ie $\sqrt{(1-p)}n \to \infty$.

So we require either $np \to \infty$ or $\sqrt{(1-p)}n \to \infty$.

marking: 2 marks for applying Chebyshev (that specific version) to the variance bound from (c). Then 3 marks for all the details of the conditions on $p = p(n)$, as long as everything is well-justified. Don't need to penalise if they miss the alternative condition ($\sqrt{(1-p)}n \to \infty$).

4. Final question asks the students to show that the Contingency tables chain is ergodic with stationary distribution being uniform. This question is broken up into a number of steps.

(a) The first task is to verify the claimed set of is the set of contingency tables for row and column sums a, b. *[5 marks]*

The first thing we observe is that $i + (a_1 - i)$ equals a_1 and $(b_1 - i) + i + (b_2 - a_1) = b_1 + b_2 - a_1$, which is $a_1 + a_2 - a_1$ (as the sum of rows must equal sum of columns), which is a_2. So both row sums are correct.

Also $i + (b_1 - i) = b_1$ and $(a_1 - i) + i + (b_2 - a_1) = b_2$, so the column sums are correct.

Hence every such table in the definition of $\Sigma_{a,b}$ satisfies the row and column sums. We have $i \geq 0$ by definition, and we have $(a_1 - i) \geq a_1 - \min\{a_1, b_1\}$, which is at least 0. So both of the row-1 entries are non-negative. Also, looking at row-2, $b_1 - i \geq b_1 - \min\{a_1, b_i\}$, which is ≥ 0. And $i + (b_2 - a_1) \geq (a_1 - b_2)_(b_2 - a_1) = 0$. So both row-2 entries are positive.

Hence all the claimed mini-tables are feasible

Finally we need to show that these are the *only* feasible tables to satisfy row sums (a_1, a_2) and column sums (b_1, b_2).

If $i > a_1$ then the top-right cell becomes negative, and if $i > b_1$ the bottom-left cell would be negative. So the possibility of a solution with $i > \min\{a_1, b_1\}$ is ruled out. Also, we certainly need $i \geq 0$. So the only question is whether some solution with $0 \leq i < a_1 - b_2$ might exist. Well, if we did have $i < a_1 - b_2$, then the top-left cell would satisfy $(a_1 - i) > b_2$. This would mean that the bottom-left cell would need to be strictly negative to satisfy column sum b_2, which is not allowed.

So yes, the defined set for $\Sigma_{a,b}$ is exactly the set of all feasible mini-tables satisfying the row sums a and column sums b.

Marking (5marks): 2 marks for verifying that the row and column sums match. 1 mark for showing that the defined entries are always positive within the bounds of the range for i. 2 marks for showing any i out-of-range would induce a negative cell somewhere in the table, hence these are the only tables. They might do the last two parts together, of course.

(b) We are given two tables $X, Y \in \Sigma_{r,c}$ and need to find a sequence of Markov chain transitions from X to Y. *[10 marks]*

We can assume $X \neq Y$ (otherwise we are done).

Consider the table $X - Y$ (subtraction carried out in a pointwise fashion), which must have row and column sums 0 (since X and Y) had identical row and column sums). By $X \neq Y$, there is at least one cell (i, j) of the table such that $(X - Y)_{ij} \neq 0$. Also note that this could be strictly positive or strictly negative, but regardless of which there is at least one cell in the same column of the opposite sign (to give overall column sum 0). Hence we can assume without loss generality that our cell of focus (i, j) has $(X - Y)_{ij} < 0$, meaning $X_{ij} < Y_{ij}$.

Now because row i of $X-Y$ must have sum 0, there is also some column index k, $k \neq j$ such that $(X-Y)_{i,k} > 0$. Also because column j of $X-Y$ must have sum 0, there is some row index h, $h \neq i$ such that $(X-Y)_{h,j} > 0$.

Now set $\widehat{\imath} = \min\{|(X-Y)_{ij}|, |(X-Y)_{ik}|, |(X-Y)_{hj}|\}$. This is definitely at least 1. Consider the following "mini-table" for the rows i, h and columns j, k

$$\begin{bmatrix} X_{ij} + \widehat{\imath} & X_{ik} - \widehat{\imath} \\ X_{hj} - \widehat{\imath} & X_{hk} + \widehat{\imath} \end{bmatrix}.$$

It is clear that replacing the $(i, h) \times (j, k)$ submatrix with the table above will preserve the row and column sums of X. However, we must verify that this is a legitimate element of $\Sigma_{a,b}$ for $a_1 = X_{ij} + X_{ik}, a_2 = X_{hj} + X_{hk}$ and $b_1 = X_{ij} + X_{hj}, b_2 = X_{ik} + X_{hk}$. All we need to verify is that all four entries are nonnegative. For the top-left and bottom right, this is obviously true by definition, given that $\widehat{\imath}$ is strictly positive and all the entries of X were non-negative. For the top-right, we use the fact that $X_{ik} - \widehat{\imath} \geq X_{ik} - |X_{ik} - Y_{ik}|$, and by the negativity of $(X - Y)_{ik}$, we have $-|X_{ik} - Y_{ik}| = X_{ik} - Y_{ik}$. So we have $X_{ik} - \widehat{\imath} \geq X_{ik} - (X_{ik} - Y_{ik}) = Y_{ik}$, and Y_{ik} is non-negative. A similar argument shows $X_{hj} - \widehat{\imath}$ is non-negative.

So this is a legitimate replacement subtable for rows i, h and columns j, k, hence there is some (strictly) positive probability of this transition being taken $(\frac{2}{m(m-1)} \frac{2}{n(n-1)} \frac{1}{|\Sigma_{a,b}|})$. Now we refer to the metric d.

Base case: Suppose we had $d(X, Y) > 0$ but that the *only* cells which differ in X and Y are $(i, j), (i, k), (h, j), (h, k)$. Then we know that $-(X_{ij} - Y_{ij}) = -(X_{hk} - Y_{hk}) = (X_{ik} - Y_{ik}) = (X_{hj} - Y_{hj})$. Hence $\widehat{\imath}$ will attain the value $|(X_{ij} - Y_{ij})|$, and the replacement subtable above will transform X into Y in one step.

Note also that this subcase contains the case when $d(X, Y) = 4$ (and note that the minimum for $d(.,.)$ is 4, and that d is always even).

General Case: Suppose now that there are more than 4 cells that differ between X and Y. Then we carry out the transition on $\{i, h\} \times \{j, k\}$ on X as described above to obtain Z^1. Now consider the value of $d(Z^1, Y)$, and note that

$$\begin{aligned} d(Z^1, Y) &= d(X, Y) - |X_{ij} - Y_{ij}| - |X_{hj} - Y_{hj}| - |X_{ik} - Y_{ik}| - |X_{hk} - Y_{hk}| \\ &\quad + |X_{ij} + \widehat{\imath} - Y_{ij}| + |X_{hj} - \widehat{\imath} - Y_{hj}| + |X_{ik} - \widehat{\imath} - Y_{ik}| + |X_{hk} + \widehat{\imath} - Y_{hk}| \end{aligned}$$

Note that $X_{ij} - Y_{ij}$ was negative, and $\widehat{\imath}$ is positive by at most $|X_{ij} - Y_{ij}|$, hence $|X_{ij} + \widehat{\imath} - Y_{ij}| = |X_{ij} - Y_{ij}| - \widehat{\imath}$.

Also by positivity of $X_{ik} - Y_{ik}$ and $-\widehat{\imath}$ being negative by at most $|X_{ik} - Y_{ik}|$, hence $|X_{ik} + \widehat{\imath} - Y_{ik}| = |X_{ik} - Y_{ik}| - \widehat{\imath}$. We have a similar argument for the (h, j) cell.

Hence

$$d(Z^1, Y) = d(X, Y) - 3\widehat{\imath} - |X_{hk} - Y_{hk}| + |X_{hk} + \widehat{\imath} - Y_{hk}|$$

In the case of the (h,k) cell, it is possible that $|X_{hk} + \widehat{\iota} - Y_{hk}|$ can be greater than $|X_{hk} - Y_{hk}|$, but this increase will be by at most $\widehat{\iota}$. Hence we are guaranteed that

$$d(Z^1, Y) \quad \leq \quad d(X, Y) - 2\widehat{\iota},$$

and by $\widehat{\iota} \geq 1$, we have made a strict reduction in the value of the metric.

Hence by the fact that $d(X, Y)$ is finite, and that we reduce the value of the metric by at least 2 in each (tailored) transition, we only need to construct a sequence of finite length to move from X to Y.

Marking (10 marks): They must identify the *specific* transition which will reduce the value of d. 3 marks for describing how we choose (i, j) and (i, k) and (j, h), and then 2 marks for explaining exactly how to alter the values of the 4 cells in X (they might forget about the "min" and instead just change by 1, this is fine). 2 marks for all the details of why this is a legitimate substitution for the 2×2 matrix (and they lose these if they start with (i, j) where $X_{ij} - Y_{ij}$ is positive, as we won't know whether we can reduce X_{hk}). 1 mark for remembering to tell us the probability will be positive. The remaining 2 marks for a bit of an induction argument (don't need something laboured, just observation about the base case, and that if not in base case, that we reduce d).

(c) We have already shown that for every $X \neq Y$, that we have a sequence of contingency tables connecting X to Y, of some length $\ell \geq 1$ (ℓ being the number of transitions along this path). *[10 marks]*

Now we make an observation that for any X and any rows $i, i', i \neq i'$ and columns $j, j', j \neq j'$, that the set of subtables in $\Sigma_{a,b}$ includes the *actual values* that X takes on these four cells (all of X_{ij}, X_{ik}, X_{hj} and X_{hk} are non-negative, and have the row sums a and column sums b). Hence, having already selected i, i' and j, j', there is a probability $\frac{1}{|\Sigma_{a,b}|}$ of keeping the values exactly as they were. Note that this will be an option regardless of how the rows were chosen and the (weighted by $\frac{2}{m(m-1)} \frac{2}{n(n-1)}$) probabilities will add up to a reasonable value. But all we need to know is that there is some strictly positive probability that X does not change, ie $M[X, X] > 0$ for any $X \in \Sigma_{r,c}$.

This tells us that for any sequence of states of length ℓ from X to Y, there is also a sequence of length $\ell + 1$ (just add a self-loop at any of the intermediate states). And for $\ell \geq 1$, the $\gcd\{\ell, \ell + 1\} = 1$. Hence the chain is *aperiodic*, as requested.

Marking (10 marks): 6 marks for stating that the chain has a positive probability of staying where it is, for any X, and for taking this fact to show we have $\ell + 1$ as well as ℓ, giving $\gcd$ 1. The other 4 marks for the details of why we have a self-loop (referring to the details of how transitions are taken). Lose a mark or two if their arguments are scanty.

(d) We now are asked to show that the chain has the uniform stationary distribution. First note that by (b) and (c) there is a *unique* stationary distribution, hence if we can show that the uniform distribution satisfies the stationarity conditions, it is that distribution. *[5 marks]*

We need to show that for every $X \in \Sigma_{r,c}$, that

$$\sum_{Y \in \Sigma_{r,c}} \frac{1}{|\Sigma_{r,c}|} M[Y, X] \stackrel{?}{=} \frac{1}{|\Sigma_{r,c}|},$$

where the $\frac{1}{|\Sigma_{r,c}|}$ represent $\pi(Y), \pi(X)$ respectively. This will be true *if and only if* we have

$$\sum_{Y \in \Sigma_{r,c}} M[Y, X] \stackrel{?}{=} 1.$$

Now consider the details of the transition from X to Y, assuming the two rows i, i' and columns j, j' have already been chosen. In making this transition, the values of the subtable change, but a_1, a_2, b_1, b_2 do not. And all the cells outside that table are identical in X and Y. Hence the set of available subtables $\Sigma_{a,b}$ is exactly the same regardless of whether we are considering i, i', j, j' in X or in Y, and hence the probability of making either transition ($X \to Y$ or $Y \to X$) is the same. Hence (assuming $X \neq Y$) we know $M[X, Y] = M[Y, X]$. For $X = Y$ it is certainly true that $M[X, Y] = M[X, X] = M[Y, X]$. Hence we can substitute $M[X, Y]$ for $M[Y, X]$ to derive the equivalent condition

$$\sum_{Y \in \Sigma_{r,c}} M[X, Y] = 1,$$

which is always true.

Hence the stationary distribution is uniform.

Marking (5 marks): They have to have written some explanations. Just a sequence of equations is not enough. 1 mark for discussing the fact we only need to verify, 1 mark for writing it in terms of the $\sum_Y M[Y, X] = 1$, 1 marks for justifying *why* $M[Y, X] = M[X, Y]$ for all X, Y, and 2 marks for finishing off. Especially important they have written some comments like "this is equivalent to showing" or "it is enough to show" to keep reminding us that this is what we *want*, not what we have (unless they happen to work forwards from $\sum_Y M[X, Y]$, which is very unlikely).