1. (a) i.

Solution: Let $n \in \mathbb{N}$, and for each $i \in \{1, \ldots, n\}$ let X_i be the indicator variable that that voucher i is redeemed (that is $X_i = 1$ if voucher i is redeemed and $X_i = 0$ otherwise). We assume that the event of a voucher being redeemed is independent of the event of any other voucher being redeemed. For each $i \in \{1, \ldots, n\}$,

$$\mathbb{E}[X_i] = 1 \cdot \Pr[X_i = 1] + 0 \cdot \Pr[X_i = 0] = p = \tfrac{4}{5}.$$

We let Y be the incurred cost of the distribution voucher, that is,

$$Y = \sum_{i=1}^{n} \left(\frac{1}{2} X_i \right) = \frac{1}{2} \sum_{i=1}^{n} X_i$$

as each voucher costs £0.50. See that

$$\mathbb{E}[Y] = \mathbb{E}\left[\frac{1}{2} \sum_{i=1}^{n} X_i \right] = \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{np}{2}.$$

Thus, by Markov's inequality,

$$\Pr[Y \geq 600] \leq \frac{\mathbb{E}[Y]}{600} = \frac{np}{1200}.$$

Then

$$\begin{aligned} \frac{np}{1200} &\leq \frac{1}{50} \\ n &\leq \frac{1200}{50p} \\ &= 30 \end{aligned}$$

thus a greatest integer value n can take (using Markov's inequality) such that the campaign will cost Lucy at most £600 is 30.

ii.

Solution: We take X_i and Y as in the previous answer. To apply Chebyshev's inequality, we must compute the variance of Y. Let $X = \sum_{i=1}^n X_i$ (so $Y = \frac{1}{2}X$). Then

$$\begin{aligned}
\mathrm{Var}[X] &= \sum_{i=1}^n \mathrm{Var}[X_i] \\
&= \sum_{i=1}^n \left(\mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2\right) \\
&= \sum_{i=1}^n \left(p - p^2\right) \\
&= np(1-p)
\end{aligned}$$

by the linearity of variance for pairwise independence, and that X_i is an indicator variable (so $\mathbb{E}[X_i^2] = \mathbb{E}[X_i]$).

Lemma. *Let (Ω, Pr) be a finite discrete probability space with random variable Z and $a \in \mathbb{R}$. Then*

$$\mathrm{Var}[aZ] = a^2 \, \mathrm{Var}[Z].$$

Proof.

$$\begin{aligned}
\mathrm{Var}[aX] &= \mathbb{E}[(aX)^2] - (\mathbb{E}[aX])^2 \\
&= a^2(\mathbb{E}[X^2] - (\mathbb{E}[X])^2) \\
&= a^2 \, \mathrm{Var}[X].
\end{aligned}$$

□

Thus

$$\mathrm{Var}[Y] = \mathrm{Var}[\tfrac{1}{2}X] = \tfrac{1}{4}\mathrm{Var}[X] = \tfrac{1}{4}np(1-p).$$

See that

$$\begin{aligned}
\Pr[Y \geq 600] &= \Pr\left[Y - \tfrac{1}{2}pn \geq 600 - \tfrac{1}{2}pn\right] \\
&= \Pr\left[Y - \mathbb{E}[Y] \geq 600 - \tfrac{1}{2}pn\right] \\
&\leq \Pr\left[|Y - \mathbb{E}[Y]| \geq 600 - \tfrac{1}{2}pn\right].
\end{aligned}$$

Thus was can apply Chebyshev's inequality, but only when $600 - \frac{1}{2}pn \geq 0$; that is, $n \leq 1500$. We keep this in mind when we get to maximising n. By Chebyshev's inequality (assuming $n \leq 1500$),

$$\begin{aligned}
\Pr\left[|Y - \mathbb{E}[Y]| \geq 600 - \tfrac{1}{2}pn\right] &\leq \frac{1}{\left(600 - \frac{1}{2}pn\right)^2} \operatorname{Var}[Y] \\
&= \frac{\frac{1}{4}np(1-p)}{\left(600 - \frac{1}{2}pn\right)^2} \\
&= \frac{np(1-p)}{4\left(600 - \frac{1}{2}pn\right)^2}.
\end{aligned}$$

So

$$\begin{aligned}
\frac{np(1-p)}{4\left(600 - \frac{1}{2}pn\right)^2} &\leq \frac{1}{50}, \\
16n^2 - 48\,200n + 36\,000\,000 &\geq 0
\end{aligned}$$

which is true for

$$n \notin \left(\tfrac{25}{4}(241 - \sqrt{481}), \tfrac{25}{4}(241 + \sqrt{481})\right)$$

(as an open interval). This gives the following critical values

$$n \approx 1369.177, \qquad n \approx 1643.323$$

correct to three decimal figures. Thus $n \lessapprox 1369.177$ or $n \gtrapprox 1643.323$. Recall that $n \leq 1500$ to apply the Chebyshev's inequality, thus we take $n = 1369$. That is, a greatest integer value n can take (using Chebyshev's inequality) such that the campaign will cost Lucy at most £600 is 1369.

iii.

Solution: Recall the following variant of the Chernoff bound.

Theorem. *Let $(\Omega, \Pr)$ be a finite discrete probability space with independent random variables $\{Z_i\}_{i=1}^n$ taking values in $\{0,1\}$. Let $Z = \sum_{i=1}^n Z_i$ be their sum and $\delta \in [0, 1.81]$. Then*

$$\Pr[Z \geq (1+\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{3}\delta^2 \mathbb{E}[X]}.$$

We take $\delta = \frac{1200-np}{np}$ and see that

$$\Pr[Y \geq 600] = \Pr[Y \geq (1+\delta)\mathbb{E}[X]].$$

Thus we can apply the above Chernoff bound variant, but only when $\delta \in [0, 1.81]$, which we keep in mind. By the Chernoff bound variant stated above (assuming $\delta \in [0, 1.81]$),

$$\begin{aligned}
\Pr[Y \geq (1+\delta)\mathbb{E}[X]] &= e^{-\frac{1}{3}\delta^2 \mathbb{E}[X]} \\
&= \exp\left(-\frac{1}{3}\left(\frac{1200-np}{np}\right)^2 \left(\tfrac{1}{2}np\right)\right) \\
&= \exp\left(\frac{1200^2 - 2400np + n^2p^2}{-6np}\right).
\end{aligned}$$

So

$$\begin{aligned}
\exp\left(\frac{1200^2 - 2400np + n^2p^2}{-6np}\right) &\leq \frac{1}{50} \\
\frac{1200^2 - 2400np + n^2p^2}{-6np} &\leq -\log(50) \\
1200^2 - 2400np + n^2p^2 &\geq 6np\log(50) \\
16n^2 - (48000 + 120\log(50))n + 36000000 &\geq 0.
\end{aligned}$$

The (approximate) critical values for this inequality are

$$n \approx 1304.372, \qquad n \approx 1724.969.$$

Thus $n \lessapprox 1304.372$ or $n \gtrapprox 1724.969$. Recall that we assumed that $\delta \in [0, 1.81]$. That is,

$$\begin{aligned} \delta &\geq 0 & \delta &\leq 1.81 \\ \frac{1200 - np}{np} &\geq 0 & \tfrac{1200-np}{np} &\leq 1.81 \\ 1200 - np &\geq 0 & 1200 - np &\leq 1.81np \\ n &\leq 1500 & n &\geq \frac{1500}{2.81} \approx 533.808. \end{aligned}$$

Thus the value $n \approx 1724.969$ is not valid. So we have $n \lessapprox 1304.372$, and the greatest integer value is 1304. Therefore, a greatest integer value n can take (using a variant of the Chernoff bound) such that the campaign will cost Lucy at most £600 is 1304.

(b)

Solution: Let $S = \{W, L, D\}$ where W is the event we win, L is the event we lose, and D is the event that we draw. We let $\mathcal{X} = (X_n)_{n\in\mathbb{N}}$ be a sequence of random variables such that

$$\begin{aligned}
\Pr[X_{n+1} = W \mid X_n = W] &= \tfrac{1}{3},\\
\Pr[X_{n+1} = L \mid X_n = W] &= \tfrac{1}{3},\\
\Pr[X_{n+1} = D \mid X_n = W] &= \tfrac{1}{3},\\
\Pr[X_{n+1} = W \mid X_n = L] &= \tfrac{1}{2},\\
\Pr[X_{n+1} = L \mid X_n = L] &= 0,\\
\Pr[X_{n+1} = D \mid X_n = L] &= \tfrac{1}{2},\\
\Pr[X_{n+1} = W \mid X_n = D] &= \tfrac{1}{3},\\
\Pr[X_{n+1} = L \mid X_n = D] &= \tfrac{1}{3},\\
\Pr[X_{n+1} = D \mid X_n = D] &= \tfrac{1}{3}
\end{aligned}$$

for all $n \in \mathbb{N}$ (as in the question). For X_1, the probability of any three events happening is $\frac{1}{3}$. This is clearly a Markov chain, and also time-homogeneous (the transition probabilities depend only on the *result* of the previous, it is not a function of $n + 1$). Thus we construct the transition matrix

$$P = \begin{array}{c c} & \begin{array}{ccc} W & L & D \end{array} \\ \begin{array}{c} W \\ L \\ D \end{array} & \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \end{array}$$

of $\mathcal{X}$.

Claim. $\mathcal{X}$ is finite and ergodic. *Proof.* Consider the transitions

(i) $W \to W$,

(ii) $W \to L$,

(iii) $W \to D$,

(iv) $L \to W$,

(v) $L \to W \to L$,

(vi) $L \to D$,

(vii) $D \to W$,

(viii) $D \to L$, and

(ix) $D \to D$.

which happen with non-zero probability, this can be confirmed by inspecting P. All of these transitions have length 1, except from (v) which has length 2. But we note that $L \to W \to D \to L$ is a transition of length 3 with non-zero probability, thus the period of L must divide 2 and 3 and so it is 1. Thus $\mathcal{X}$ is *irreducible* (you can reach every state from any state with non-zero probability) and *aperiodic* (each state has period 1), thus it is ergodic. $\mathcal{X}$ is finite as $|S| = 3 < \infty$. □

As $\mathcal{X}$ is finite and ergodic, it must converge to to a unique stationary distribution. Let $\boldsymbol{\pi} = (a, b, c)$ be a stationary distribution of $\mathcal{X}$; that is, $\boldsymbol{\pi} P = \boldsymbol{\pi}$. Combining this with the fact that $\boldsymbol{\pi}$ is a distribution (so $a + b + c = 1$), we get the following system of equations.

$$\begin{aligned} \tfrac{1}{3}a + \tfrac{1}{2}b + \tfrac{1}{3}c &= a, \\ \tfrac{1}{3}a + \tfrac{1}{3}c &= b, \\ \tfrac{1}{3}a + \tfrac{1}{2}b + \tfrac{1}{3}c &= c, \\ a + b + c &= 1. \end{aligned}$$

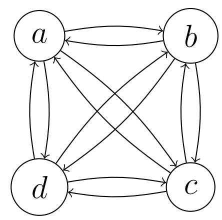
We put this into reduced echelon form, and get

$$\left(\begin{array}{ccc|c} -{}^2\!/_3 & {}^1\!/_2 & {}^1\!/_3 & 0 \\ {}^1\!/_3 & -1 & {}^1\!/_3 & 0 \\ {}^1\!/_3 & {}^1\!/_2 & -{}^2\!/_3 & \\ 1 & 1 & 1 & 1 \end{array}\right) = \left(\begin{array}{ccc|c} -4 & 3 & 2 & 0 \\ 1 & -3 & 1 & 0 \\ 2 & 3 & -4 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right)$$
$$\to \left(\begin{array}{ccc|c} 8 & 0 & 0 & 3 \\ 0 & 4 & 0 & 1 \\ 0 & 0 & 8 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

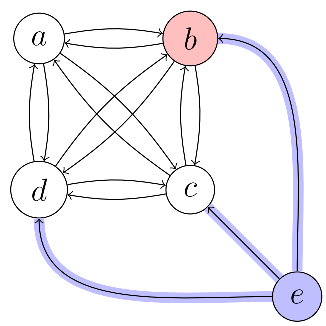
Thus $\boldsymbol{\pi} = (\frac{3}{8}, \frac{1}{4}, \frac{3}{8})$ is the *only* stationary distribution, and $\mathcal{X}$ converges to $\boldsymbol{\pi}$. This implies that, in the long run, the proportion of games that are won are $\frac{3}{8}$, the proportion that are lost is $\frac{1}{4}$.

(c) i.

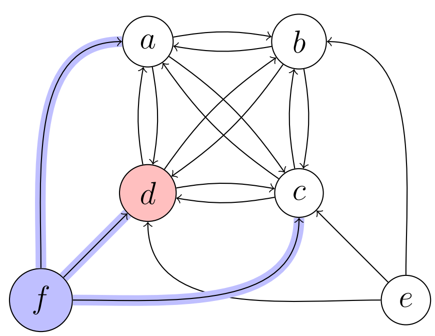
Solution: We start with the following seed graph.



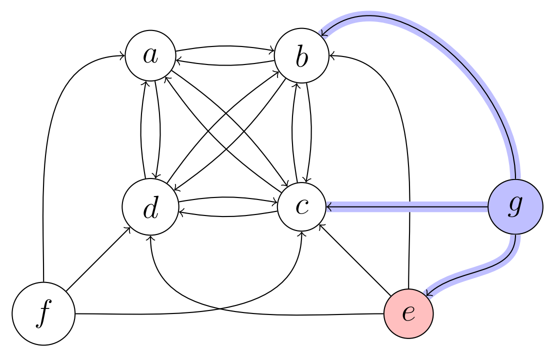
For the first step, we let $v = e$ be the new vertex and we (randomly) pick $u = b$. We sample $(w_1, w_2) = (d, c)$ from the neighbourhood of u, and thus we get the following graph (the selected vertex is coloured red, and the new vertex and edges are coloured blue).



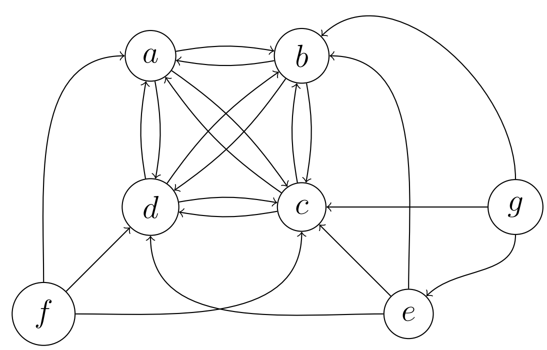
For the second step, we let $v = f$ be the new vertex and we (randomly) pick $u = d$, and sample neighbours $(w_1, w_2) = (a, c)$. Thus we get the following graph.



For the final step, we let $v = g$ be the new vertex and randomly pick $u = e$, with neighbourhood sample $(w_1, w_2) = (b, c)$. Thus we get the following graph.



Giving us the following final graph.



ii.

Solution: Denote the in-degree of a vertex $v \in V(G)$ of some graph G by $\mathrm{ideg}(v)$.
Fix i and at t fix two outcomes of (R_t, S_t):

(i) let u be the vertex we selected in the first trial, and $w_1, \ldots, w_{k-1}$ be the sampled neighbours from the first trial;

(ii) let u' be the vertex we selected in the second trial, and $w'_1, \ldots, w'_{k-1}$ be the sampled neighbours from the second trial.

Let $A = \{u, w_1, w_2, \ldots, w_{k-1}\}$, $A_j = \{a \in A : \mathrm{ideg}(a) = j\}$, $B = \{u', w'_1, w'_2, \ldots, w'_{k-1}\}$, and $B_j = \{b \in B : \mathrm{ideg}(j)\}$. $X_i^{(t)}$ is the number of nodes with in-degree i at t. Let X be the random variable $X_i^{(t)}$ given the first outcome of (R_t, S_t), and Y be the random variable $X_i^{(t)}$ given by the second outcome of (R_t, S_t).

Claim.

$$X = X_i^{(t-1)} + A_{i-1} - A_i,$$
$$Y = X_i^{(t-1)} + B_{i-1} - B_i.$$

Proof. We consider the case only for A, as it is the same for A' (and in fact any outcome). In the transition $X_i^{(t-1)} \to X_i^{(t)}$, we introduced a new vertex to the graph and connected it to each element of A. Thus the in-degree of each element of A increases by 1 (from time $t-1$ to t). As these are the only edges added, the in-degree of every node outside of A does not change. We add the nodes with in-degree i at the time $t-1$, and add or remove the nodes for which this change. Any node that had in-degree $i-1$ before this transition now has in-degree i, thus we add this. Similarly, any node that had in-degree i before now has in-degree $i+1$, so we minus this. $\square$

For the two trials we selected, the maximum difference is

$$\begin{aligned}
|X - Y| &= \left| \left(X_i^{(t-1)} + A_{i-1} - A_i \right) - \left(X_i^{(t-1)} + B_{i-1} - B_i \right) \right| \\
&= |(A_{i-1} - A_i) - (B_{i-1} - B_i)| \\
&\leq |A_{i-1} - A_i| + |B_{i-1} - B_i|
\end{aligned}$$

by the triangle inequality. But for all j, $A_j \subset A$. Thus $|A_j| \leq |A| = k$. As $|A_j| \geq 0$ for all j, we have $|A_{i-1} - A_i| \leq k$ and similarly $|B_{i-1} - B_i| \leq k$. Thus

$$|X - Y| \leq k + k = 2k$$

as required.

(d)

Solution: We construct $H \subset G$ as follows. Let $V_H = A \cup B$. For each vertex $u \in V$, flips a coin. If the coin reads tails, put u in A. If the coin reads heads, put u in A. Thus, for all $u \in V$, $\Pr[u \in A] = \frac{1}{2}$ and $\Pr[u \in B] = \frac{1}{2}$, and $A \cap B = \varnothing$. For each $uv \in E$, we let $uv \in E_H$ if and only if:

(i) $u \in A$ and $v \in B$; or

(ii) $u \in B$ and $v \in A$.

Then, for all $uv \in E$

$$\Pr[uv \in E_H] = 1 - \Pr[(u, v \in A) \vee (u, v \in B)].$$

Both events $[u, v \in A]$ and $[u, v \in B]$ are mutually exclusive (that is, they can't both happen), thus

$$\Pr[uv \in E_H] = 1 - (\Pr[u, v \in A] + \Pr[u, v \in B]).$$

The events $[u \in A]$ and $[v \in A]$ are independent, thus we conclude

$$\begin{aligned}\Pr[uv \in E_H] &= 1 - (\Pr[u \in A]\Pr[v \in A] + \Pr[u \in B]\Pr[v \in B]) \\ &= \frac{1}{2}.\end{aligned}$$

Enumerate the edges of G as $E = \{e_1, \ldots e_m\}$ and for all $i \in \{1, \ldots, m\}$ let X_i be the indicator function for edge e_i being in H. We have

$$\mathbb{E}[X_i] = 1 \cdot \Pr[e_i \in H] + 0 \cdot \Pr[e_i \notin H] = \tfrac{1}{2}.$$

Let $X = \sum_{i=1}^{m} X_i$; that is, X is the random variable corresponding to the number of edges in H in our construction. Then

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^{m} X_i\right] \\ &= \sum_{i=1}^{m} \mathbb{E}[X_i] \\ &= \frac{m}{2}.\end{aligned}$$

As the expected number of edges in our random construction H is $\frac{m}{2}$, there must be at least one H with $\frac{m}{2}$ edges.