SC9 Probability on Graphs and Lattices: Sheet #1

Due on October 22, 2025 at 12:00pm

Professor C. Goldschmidt and Professor J. Jorritsma

Ray Tsai

Problem 1

Given a finite, connected graph G, and a path of neighbouring vertices $(x_0, x_1, x_2, \dots)$ such that every vertex $v \in V(G)$ appears in the path, let $\tau_v := \inf\{n : x_n = v\}$. Let T be the subgraph of G with $V(T) = V(G)$ and edge-set

$$E(T) := \{\{x_{\tau_v - 1}, x_{\tau_v}\} : v \in V(G) \setminus \{x_0\}\}.$$

Prove that T is a spanning tree for G.

Proof. Reorder the vertices of G by their hitting times, i.e. $i < j$ if and only if $\tau_{v_i} < \tau_{v_j}$. Let $T_n \subseteq G$ be graph with edges

$$E(T_n) = \{\{x_{\tau_{v_i} - 1}, x_{\tau_{v_i}}\} : i = 0, \dots, n-1\}.$$

We prove by induction on n that T is a connected graph on $v_0, \dots, v_{n-1}$. When $n = 0$, T_0 is simply the singleton graph on v_0. Suppose $n \geq 1$. Then

$$E(T_n) = \{\{x_{\tau_{v_{n-1}} - 1}, x_{\tau_{v_{n-1}}}\}\} \cup E(T_{n-1}).$$

By induction, T_{n-1} is a connected graph on $v_0, \dots, v_{n-2}$. But then $x_{\tau_{v_{n-1}} - 1} = v_k$ for some $k < n-1$, and thus T_n is a connected graph on $v_0, \dots, v_{n-1}$. It now follows from $|E(T_n)| = |E(T)| = |V(G)| - 1$ that T is a spanning tree of G. $\square$

Problem 2

(a) Consider the *coupon collector's problem*: boxes of a certain cereal come with one of n distinct coupons, chosen uniformly at random, and you wish to collect the full set of n coupons. Show that the expected number N_n of boxes of cereal that you have to buy is such that

$$\mathbb{E}[N_n] \sim n \log n$$

as $n \to \infty$.

Proof. Let T_i be number of boxes it takes to get the i-th distinct coupon after getting the previous $i-1$ distinct coupons. Then $N_n = \sum_{i=1}^n T_i$. Note that after getting $i-1$ distinct coupons, the probability of getting a new coupon is $(n-i+1)/n$. Thus, T_i is a geometric random variable with parameter $(n-i+1)/n$, so $\mathbb{E}[T_i] = n/(n-i+1)$. By the linearity of expectation,

$$\mathbb{E}[N_n] = \sum_{i=1}^n \mathbb{E}[T_i] = \sum_{i=1}^n \frac{n}{n-i+1} = n \sum_{j=1}^n \frac{1}{j} \sim n \log n.$$

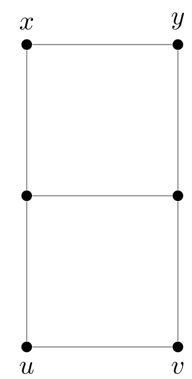
□

(b) Use (a) to give an upper bound on the expected number of steps taken by the Aldous-Broder algorithm on the complete graph K_n.

Proof. Consider the vertices of K_n as coupons. Then performing the Aldous-Broder algorithm on K_n is equivalent to collecting the full set of n coupons, except we do not get the same coupon twice in a row as there are no loops in K_n. In particular, the probability of getting a new coupon after getting $i-1$ distinct coupons is $(n-i+1)/(n-1) > (n-i+1)/n$. Thus, the expected number of steps taken by the Aldous-Broder algorithm on K_n is at most the expected number of steps taken by the coupon collector's problem on K_n, which is $n \log n$. □

Problem 3

Consider the graph G:



Let $e = \{u, v\}$ and $e' = \{x, y\}$. Let T be a UST of G. Show that $\mathbb{P}_u^G(\tau_v < \tau_u^+) = 15/22$ and $\mathbb{P}_u^{G/e'}(\tau_v < \tau_u^+) = 11/16$. Deduce that, in this case, we indeed have

$$\mathbb{P}(e \in E(T) | e' \in E(T)) \leq \mathbb{P}(e \in E(T)).$$

This question is partly intended as a reminder of how to do hitting probability calculations! You may find it helpful in each case to write down a system of simultaneous equations and solve them (using a computer if you like) to find the desired probabilities.

Proof. Let the middle left vertex be a and the middle right vertex be b. Let $P_w = \mathbb{P}_w^G(\tau_v < \tau_u)$ for $w \in V(G) \backslash \{u, v\}$. Let $P_u = \mathbb{P}_u^G(\tau_v < \tau_u^+)$. We then have

$$\begin{cases} P_u = \frac{1}{2} + \frac{1}{2} P_a \\ P_a = \frac{1}{3} P_b + \frac{1}{3} P_x \\ P_x = \frac{1}{2} P_a + \frac{1}{2} P_y \\ P_b = \frac{1}{3} + \frac{1}{3} P_a + \frac{1}{3} P_y \\ P_y = \frac{1}{2} P_b + \frac{1}{2} P_x \end{cases}$$

Solving this yields $P_u = 15/22$. For $w \in V(G) \backslash \{v\}$, define P'_w in a similar way but on G/e'. We then have

$$\begin{cases} P_u = \frac{1}{2} + \frac{1}{2} P_a \\ P_a = \frac{1}{3} P_b + \frac{1}{3} P_x \\ P_x = P_a \\ P_b = \frac{1}{3} + \frac{1}{3} P_a + \frac{1}{3} P_y \\ P_y = P_b \end{cases}$$

Solving this yields $P_u = 11/16$. □

Problem 4

For every vertex $v_i \in V(G) \setminus \{v_0\}$, select a directed edge $v_i \vec{w}_i$. Prove that this collection of directed edges is either a spanning tree on G directed towards v_0, or includes a directed cycle.

Proof. Suppose this collection of directed edges does not contain a directed cycle. Since we are selecting a directed edge pointing outwards from each vertex $v_i \in V(G) \setminus \{v_0\}$, none of the vertices v_i can be a sink in the resulting directed graph T. But then v_0 is the only possible sink in T. Since T is a DAG, it must contain a sink, so T is a DAG with a single sink at v_0, which makes T a spanning tree. $\square$

Problem 5

By reference to Wilson's algorithm, or otherwise, prove that in a finite or recurrent connected graph G, the law of the loop-erased random walk path from x to y is the same as the law of the loop-erased random walk path from y to x.

Note: the loop-erased random walk path from x to y is constructed by taking the (almost surely finite) path of a random walk from x stopped at τ_y, and then loop-erasing it.

Proof. Since G is connected and recurrent, we may generate FUSF on G using Wilson's algorithm. For $v \in V(G)$, let T_v be a random variable for the FUSF on G generated by Wilson's algorithm by setting $v_0 = v$. By Proposition 1.19, T_v is a.s. connected. Thus there exists a path a.s. from any $u \in V(G)\backslash\{v\}$ to v, denoted $P_v(u)$, and we note that $P_v(u)$ is a LERW. Since the distribution of the generated FUSF is independent of the root, μ_{T_v} is the same for any $v \in V(G)$. Let $A \subset E(G)$ denote a collection of edges that forms a path between x and y. SInce $\mu_{T_x} = \mu_{T_y}$,

$$\mathbb{P}(P_x(y) = A) = \mathbb{P}(A \subset E(T_x)) = \mathbb{P}(A \subset E(T_y)) = \mathbb{P}(P_y(x) = A).$$

This completes the proof. $\square$

Problem 6

Let T_n be a UST of the complete graph K_n. Let v, w, w' be distinct vertices in K_n and consider edges $e = \{v, w\}$ and $e' = \{v, w'\}$. Use the Aldous-Broder algorithm to prove that e, e' are negatively associated in T_n i.e. that

$$\mathbb{P}(e, e' \in E(T_n)) \leq \mathbb{P}(e \in E(T_n))\mathbb{P}(e' \in E(T_n))$$

for all sufficiently large n.

Proof. Consider the Aldous-Broder algorithm starting from v. Let A be the event that $e \in E(T_n)$ and B be the event that $e' \in E(T_n)$. Let X_i denote the i-th step of the SRW from v. Note that $e \in E(T_n)$ if and only if e is the first entry into w. That is, either it happnes immediately, or requires the SRW to return to v before reaching w. Thus by the Strong Markov Property,

$$\mathbb{P}(A) = \frac{1}{n-1} + \mathbb{P}_v(\tau_v^+ < \tau_w)\mathbb{P}(A),$$

where $\mathbb{P}_v$ denotes the probability measure on the SRW starting from v. Since $\mathbb{P}_v(\tau_v^+ = \tau_w) = 0$, rearranging the above equation yields

$$\mathbb{P}(A) = \frac{1}{(n-1)\mathbb{P}(\tau_v^+ > \tau_w)}.$$

Note that

$$\begin{aligned}\mathbb{P}_v(\tau_v^+ > \tau_w) &= \mathbb{P}_v(\tau_v^+ > \tau_w \mid X_1 = w)\mathbb{P}_v(X_1 = w) + \sum_{u \in V(G)\setminus e} \mathbb{P}_v(\tau_v^+ > \tau_w \mid X_1 = u)\mathbb{P}_v(X_1 = u) \\ &= \frac{1}{n-1} + \frac{1}{n-1} \sum_{u \in V(G)\setminus e} \mathbb{P}_u(\tau_v > \tau_w).\end{aligned}$$

By symmetry, $\mathbb{P}_u(\tau_v > \tau_w) = \mathbb{P}_u(\tau_w > \tau_v)$, so $\mathbb{P}_u(\tau_v > \tau_w) = 1/2$. Hence, we have

$$\mathbb{P}_v(\tau_v^+ > \tau_w) = \frac{1}{n-1} + \frac{n-2}{n-1} \cdot \frac{1}{2} = \frac{n}{2(n-1)},$$

and thus $\mathbb{P}(A) = 2/n$. By symmetry, $\mathbb{P}(B) = 2/n$.

We now compute $\mathbb{P}(A \cap B)$. We may write

$$\mathbb{P}(A \cap B) = \frac{1}{n-1} \cdot \sum_{u \in V(G)\setminus\{v\}} \mathbb{P}(A \cap B \mid X_1 = u)$$

If $X_1 = w$, then by the same symmetry argument,

$$\mathbb{P}(A \cap B \mid X_1 = w) = \mathbb{P}(B|X_1 = w) = \mathbb{P}_w(\tau_v < \tau_{w'})P(B) = \frac{1}{2} \cdot \frac{2}{n} = \frac{1}{n}.$$

Similarly, we also have $\mathbb{P}(A \cap B \mid X_1 = w') = \frac{1}{n}$. Now suppose $X_1 = u$ for some $u \in V(G)\setminus\{v, w, w'\}$. Then,

$$\mathbb{P}(A \cap B \mid X_1 = u) = \mathbb{P}_u(\tau_v < \tau_w \text{ and } \tau_v < \tau_{w'})\mathbb{P}(A \cap B) = \frac{1}{3} \cdot \mathbb{P}(A \cap B),$$

as the probability of first reaching either v, w, or w' is the same. Substituting back to the initial equation,

$$\mathbb{P}(A \cap B) = \frac{1}{n-1}\left(\frac{2}{n} + (n-3) \cdot \frac{1}{3} \cdot \mathbb{P}(A \cap B)\right).$$

Rearranging yields

$$\mathbb{P}(A \cap B) = \frac{3}{n^2} \leq \frac{4}{n^2} = \mathbb{P}(A)\mathbb{P}(B).$$

This completes the proof. □

Problem 7

Prove that the free uniform spanning forest on an infinite, connected, locally-finite graph G has no finite components almost surely.

Proof. Let (G_n) be some exhaustion of G, with associated USTs (T_n). Let F be a FUSF of G. Let $C \subset V(G)$ be a finite set of vertices, and define

$$\mathcal{K}_C = \{\{u,v\} \in E(G) : u \in C, v \in V(G)\backslash C\}.$$

Since G is conenected and locally finite, $\mathcal{K}_C$ is nonempty and finite. Let E_C be the event that C is a component in F. Then E_C implies the cylinder event $A_C = \{E(F) \cap \mathcal{K}_C = \emptyset\}$, so $\mathbb{P}(E_C) \leq \mathbb{P}(A_C)$. We now show that $\mathbb{P}(A_C) = 0$. Note that

$$\mathbb{P}(A_C) = \mu^F(A_C) = \lim_{n\to\infty} \mu_{T_n}(A_C) = \lim_{n\to\infty} \mathbb{P}(E(T_n) \cap \mathcal{K}_C = \emptyset).$$

Suppose n is large enough such that C is strictly contained in $V(G_n)$. Since G_n is connected and T_n is a spanning tree, T_n must contain a path from C to $V(G_n)\backslash C$. That is, T_n must contain some edge in G. But then

$$\mathbb{P}(A_C) = \lim_{n\to\infty} \mathbb{P}(E(T_n) \cap \mathcal{K}_C = \emptyset) = 0.$$

It now follows that

$$\mathbb{P}(F \text{ has some finite component}) = \mathbb{P}\left(\bigcup_{\substack{C\subset V(G)\\ |C|<\infty}} E_C\right) \leq \sum_{\substack{C\subset V(G)\\ |C|<\infty}} \mathbb{P}(E_C) \leq \sum_{\substack{C\subset V(G)\\ |C|<\infty}} \mathbb{P}(A_C) = 0.$$

□

Problem 8

Let G be the lattice $\mathbb{Z}^2$. Given the box $G_n = [-n, n]^2 \cap \mathbb{Z}^2$, the Dobrushin wiring G_n^{Dob} consists of adding a vertex u_n, and an edge between u_n and each of the $4n+2$ vertices which lie either on the left-boundary or the right-boundary. Let T_n^{Dob} be the UST on G_n^{Dob}, and let $\mu_{G_n^{\text{Dob}}}$ be the probability measure on Ω_G describing the restriction of T_n^{Dob} to $G_n \subset G$. Prove that $\mu_{G_n^{\text{Dob}}} \Rightarrow \mu^F$.

Proof. Let $A \subset E(G)$ be some finite set of edges, and let $\mathcal{C}_A = \{\omega \in \Omega_G : w(e) = 1, \forall e \in A\}$. Since $\mathcal{C}_A$ is an increasing cylinder event, by Proposition 1.17, it suffices to show that

$$\lim_{n \to \infty} \mu_{G_n^{\text{Dob}}}(\mathcal{C}_A) = \mu^F(\mathcal{C}_A).$$

Assume n large enough such that $A \subset E(G_n)$. Consider the wired subgraph (G_n^W) and the associated USTs (T_n^W). Notice that $G_n \subseteq G_n^{\text{Dob}} \subseteq G_n^W$, so

$$\mu_{T_n^W}(\mathcal{C}_A) \leq \mu_{G_n^{\text{Dob}}}(\mathcal{C}_A) \leq \mu_{T_n}(\mathcal{C}_A).$$

But then G is recurrent and connected, so by Proposition 1.26,

$$\lim_{n \to \infty} \mu_{T_n^W}(\mathcal{C}_A) = \lim_{n \to \infty} \mu_{T_n}(\mathcal{C}_A) = \mu^F(\mathcal{C}_A).$$

The desired result now follows from sandwiching. $\square$

Problem 9

Let G be an infinite, recurrent, connected graph with an exhaustion (G_n). By coupling a random walk on G and a random walk on G_n appropriately, show that the Aldous-Broder algorithm also generates the UST on G (which you should view as the FUSF on G, defined along an exhaustion).

Proof. Let (T_n) be the associated USTs of (G_n). Let $A \subset E(G)$ be finite, with $\mathcal{C}_A$ the corresponding cylinder event. Let $\mathcal{A} \subset V(G)$ denote the finite set of vertices incident to A. Assume that n is large enough that $\mathcal{A} \subset V_n$.

Now run Aldous-Broder algorithm on G. Since G is recurrent and connected, the SRW will hit every vertex in $\mathcal{A}$. Note that we can also run Aldous-Broder algorithm on G_n using the same SRW, and the partial subtrees generated will be the same until the SRW hits ∂G_n. But then the restrictions of T and T_n to $\mathcal{A}$ are different only if the SRW hit ∂G_n before hitting every vertex in $\mathcal{A}$. Let $\tau_{\partial G}$ denote the hitting time of ∂G and we have

$$\begin{aligned}
|\mathbb{P}(A \subset E(T)) - \mathbb{P}(A \subset E(T_n))| &\leq \mathbb{P}(\{A \subset E(T)\} \Delta \{A \subset E(T_n)\}) \\
&\leq \mathbb{P}(T \text{ restricted to } \mathcal{A} \text{ not built before } \tau_{\partial G_n}) \\
&\leq \mathbb{P}\left(\bigcup_{v \in \mathcal{A}} \{\tau_v > \tau_{\partial G_n}\} \right) \\
&\leq \sum_{v \in \mathcal{A}} \mathbb{P}(\tau_v > \tau_{\partial G_n}),
\end{aligned}$$

by the union bound. Since G is recurrent, $\mathbb{P}(\tau_v > \tau_{\partial G_n}) \to 0$ as $n \to \infty$. Thus we have

$$\lim_{n \to \infty} \mathbb{P}(A \subset E(T_n)) = \mathbb{P}(A \subset E(T)).$$

The result now follows from Proposition 1.17. □

Problem 10

Consider again the coupon collector's problem from Question 2. For $k \geq 0$ let $C_n(k)$ be the number of coupons which have not yet been collected by step k, so that $C_n(0) = n$ and $C_n(1) = n - 1$.

(a) Let $M_n(k) = \left(1 - \frac{1}{n}\right)^{-k} C_n(k)$. Show that $\mathbb{E}[M_n(k+1) | C_n(k)] = M_n(k)$ (i.e. the process $(M_n(k))_{k \geq 0}$ is a martingale).

(b) Hence show that $\mathbb{P}(N_n > \lceil n \log n + cn \rceil) \leq e^{-c}$ for any $c > 0$.

Problem 11

Let $G = (V, E)$ be a connected recurrent graph, and let $(X_n)_{n>0}$ be a simple random walk on G, which moves around on the vertices of the graph, at each step independently moving to a neighbour of its current position chosen uniformly at random.

(a) For a fixed directed edge (v, w), find the mean return time to (v, w).

(b) Deduce the edge-commute identity:

$$\mathbb{E}_v[\tau_w] + \mathbb{E}_w[\tau_v] \leq 2|E|,$$

where $\tau_u := \inf\{n \geq 0 : X_n = u\}$ for $u \in V$.

(c) Let t_{cov} be the cover time of G, that is the first time that the SRW has visited all the vertices. Prove that for any spanning tree t of G and any vertex $u \in V$ we have

$$\mathbb{E}_u[t_{\text{cov}}] \leq \sum_{\{v,w\} \in t} (\mathbb{E}_v[\tau_w] + \mathbb{E}_w[\tau_v]),$$

and deduce an upper bound on $\max_{u \in V} \mathbb{E}_u[t_{\text{cov}}]$ in terms of $|E|$.

(d) Further deduce that the expected number of steps in the Aldous-Broder algorithm is bounded above by $|V|^3$ for any graph.

(e) Give an example of a graph for which the upper bound in (c) is of the correct order and an example of a graph for which it is not.

Problem 12

On the complete graph K_n, with $n \geq 2$, Aldous (1990) gave another algorithm to generate a UST, as follows. Let $U_2, \ldots, U_n$ be uniform on $\{1, 2, \ldots, n-1\}$. Start from a single vertex labelled 1.

- For $2 \leq i \leq n$ connect vertex i to vertex $V_i = \min\{U_i, i-1\}$.
- Relabel vertices $1, \ldots, n$ as $\pi(1), \ldots, \pi(n)$ where π is a uniform random permutation of $1, \ldots, n$.

(Note that this algorithm has only $n-1$ steps, and so is considerably more efficient than Aldous-Broder on K_n!)

(a) Starting from the Aldous-Broder algorithm, or otherwise, verify that this algorithm indeed yields a UST of K_n.

(b) Let $L_n^{(1)}$ be the first index at which $\min\{U_i, i-1\} \neq i-1$. Find $\mathbb{P}(L_n^{(1)} \geq k+1)$.

(c) Show that $L_n^{(1)}/\sqrt{n} \to L^{(1)}$ where $L^{(1)}$ has density $f(x) = xe^{-x^2/2}$, $x \geq 0$.

(d) Now let $L_n^{(2)}$, $L_n^{(3)}, \ldots$ be the successive subsequent indices at which $\min\{U_i, i-1\} \neq i-1$. What can you say about the joint limit in distribution of

$$\frac{1}{\sqrt{n}}(L_n^{(1)}, L_n^{(2)} - L_n^{(1)}, \ldots, L_n^{(m)} - L_n^{(m-1)})$$

for $m \geq 2$ as $n \to \infty$?

This shows that the correct "length-scale" for the UST on K_n is $\sqrt{n}$. Indeed, much more is true: the result above is an important aspect of the convergence of the UST on K_n, on rescaling by $1/\sqrt{n}$ to the so-called Brownian continuum random tree.

Problem 13

Let T be the UST on $\mathbb{Z}^2$, and let S_n be the subgraph of T induced on the box $\Lambda_n = [-n, n]^2 \cap \mathbb{Z}^2$.

(a) Find the best constants α_n, β_n such that

$$\alpha_n \leq |E(S_n)| \leq \beta_n$$

holds almost surely. Hint: you may find it helpful to consider the connectivity of the wired version of S_n.

(b) Hence, or otherwise, show that if e is any edge of $\mathbb{Z}^2$ then

$$\mathbb{P}(e \in E(T)) = \frac{1}{2}.$$

Problem 14

Let T be the UST on $\mathbb{Z}^2$.

(a) Show that there exist two adjacent vertices on the boundary of the box $[-n, n]^2 \cap \mathbb{Z}^2$ that are connected by a path in T with length at least $2n$.

(b) Let L be the length of the path from $(0, 0)$ to $(0, 1)$ in T. Use (a) to show that

$$\mathbb{P}(L \geq 2n) \geq \frac{1}{8n}.$$