Statistics 251/551 spring 2013
Homework # 4
Due: Wednesday 20 February

If you are not able to solve a part of a problem, you can still get credit for later parts: Just assume the truth of what you were unable to prove in the earlier part.

This homework will step you through the proof of Lemma 1 in the Jerrum (1995) paper.[1] I mostly use Jerrum's notation, except that the number of colors is q, not k.

The setting
The set of available "colors" is $\mathcal{C} = \{1, 2, \ldots, q\}$. We have a graph $\mathcal{G}$ on a finite set of vertices $\mathcal{V} = \{v_1, \ldots, v_n\}$ with edge set $\mathcal{E} = \{e_1, \ldots, e_R\}$. The degree of a vertex v, denoted by $\deg(v)$, is the number of edges having v as one of the endpoints. The neighborhood of vertex v is the set of vertices connected to v by an edge,

$$\mathcal{N}(v) := \{w \in \mathcal{V} : w \neq v \text{ and } \{v, w\} \in \mathcal{E}\}.$$

A ***coloring of the graph*** is a map $\sigma : \mathcal{V} \to \mathcal{C}$. The coloring is ***proper*** if the two vertices that make up each edge are assigned different colors, that is, if

$$\sigma(v) \notin \{\sigma(w) : w \in \mathcal{N}(v)\} \qquad \text{for every } v \in \mathcal{V}.$$

Denote the set of all proper colorings of the graph by $\mathcal{S}$.

It is easy to see (by means of a greedy coloring method) that $\mathcal{S}$ is nonempty if $q > \Delta$. For $q > 2\Delta$ Jerrum's algorithm generates observations from the uniform distribution π on $\mathcal{S}$ by means of a Markov chain that converges very rapidly towards π.

Transition probabilities
The transition probabilities $P(\sigma, \tau)$ for the chain are defined implicity by a random method for producing a new coloring τ from a coloring σ.

First define a function TRY that generates a new coloring τ given a new color c, a vertex v, and a current coloring σ. Define $\tau = TRY(c, v, \sigma)$ by

(i) if $c \in \{\sigma(w) : w \in \mathcal{N}(v)\}$ then $\tau = \sigma$

(ii) if $c \notin \{\sigma(w) : w \in \mathcal{N}(v)\}$ then $\tau(v) = c$ and $\tau(w) = \sigma(w)$ for all w not equal to v.

In other words, $TRY(c, v, \sigma)$ changes the color of v to c, provided the resulting coloring is proper. If the proposed change would create an improper coloring, σ is left unchanged.

[1] Available at ***http://onlinelibrary.wiley.com/doi/10.1002/rsa.3240070205/abstract***. You might need to access the site via ***http://sfx.library.yale.edu/sfx_local/azlist*** .

Here is the random procedure corresponding to $P(\sigma, \tau)$:

(a) Choose a vertex V at random from (the uniform distribution on) $\mathcal{V}$.

(b) Choose a color C at random from (the uniform distribution on) $\mathcal{C}$.

(c) Define $\tau = TRY(C, V, \sigma)$.

The coupling

Jerrum's proof works by creating a coupling of a Markov chain $\{Y_t : t \geq 0\}$ with state space $\mathcal{S}$ and an arbitrary (but fixed) initial distribution μ and another Markov chain $\{X_t : t \geq 0\}$ with state space $\mathcal{S}$ and initial distribution π, using a method a little like the coupling used to prove the BLT.

Write info_t for the information corresponding to everything that has happened up to the completion of step t. Initially $X_0 \sim \pi$ and $Y_0 \sim \mu$, independently. After the completion of step t the X_{t+1} and Y_{t+1} values are coupled as follows

(a) Choose a new vertex V_{t+1} at random from (the uniform distribution on) $\mathcal{V}$, independently of info_t.

(b) Independently of V_{t+1} and of info_t, choose a color C at random from (the uniform distribution on) $\mathcal{C}$.

(c) Define $X_{t+1} = TRY(C_{t+1}, V_{t+1}, X_t)$. Based on info_t and V_{t+1}, construct a one-to-one function $g : \mathcal{C} \to \mathcal{C}$ then define $Y_{t+1} = TRY(g(C_{t+1}), V_{t+1}, Y_t)$.

Remark. The random color $g(C_{t+1})$ is also uniformly distributed on $\mathcal{C}$, independently of info_t and V_{t+1}. The change from Y_t to Y_{t+1} still follows the P transition probabilities; marginally, Y is still just a Markov chain with initial distribution μ and transition matrix P.

If $X_t(V_{t+1}) \neq Y_t(V_{t+1})$ then g is taken to be the identity map (that is, $g(c) = c$ for all c in $\mathcal{C}$) and $Y_{t+1} = TRY(C_{t+1}, V_{t+1}, Y_t)$. The cleverness in Jerrum's algorithm comes from the choice of g for the cases where $X_t(V_{t+1}) = Y_t(V_{t+1})$.

We need some notation for the case where X_t and Y_t agree on the color given to vertex V_{t+1}. After step t of the algorithm the vertex set $\mathcal{V}$ finds itself partitioned into two subsets: $A_t = \{v \in \mathcal{V} : X_t(v) = Y_t(v)\}$, the set of vertices where the two colorings agree; and $D_t = \{v \in \mathcal{V} : X_t(v) \neq Y_t(v)\}$, the set where they disagree.

For the sake of notational clarity let me omit some subscripts t and $t + 1$, writing V for V_{t+1} and so on. Write $K_X(V)$ for $\{X_t(w) : w \in \mathcal{N}(V)\}$, the set of colors used by X_t for the neighbors of V. Define $K_Y(V)$ analogously. Then $\mathcal{C}_X(V) := K_X(V) \cap K_Y^c(V)$ is the set of colors used only by X_t for the neighbors of V and $\mathcal{C}_Y(V) := K_Y(V) \cap K_X^c(V)$ is the set of colors used only by Y_t for the neighbors of V.

The g is chosen in a way that discourages changes that would turn an agreement at vertex V into a disagreement.

If either $\mathcal{C}_X(V)$ or $\mathcal{C}_Y(V)$ is empty let g be the identity map on $\mathcal{C}$.

If both $\mathcal{C}_X(V)$ and $\mathcal{C}_Y(V)$ are nonempty the smaller of the two sets of colors drives the construction. For concreteness suppose $0 < \#\mathcal{C}_X(V) \le \#\mathcal{C}_Y(V)$, so that the distinct colors in $\mathcal{C}_X(V)$ can be enumerated as $c_1, c_2, \ldots, c_r$ and the distinct colors in $\mathcal{C}_Y(V)$ can be enumerated as $c'_1, c'_2, \ldots, c'_{r'}$, with $r \le r'$. Define g by

$$\begin{aligned} &g(c_i) = c'_i \text{ and } g(c'_i) = c_i \qquad \text{for } i = 1, 2, \ldots, r \\ &g(c) = c \qquad \text{otherwise} \end{aligned}$$

If $\#\mathcal{C}_X(V) > \#\mathcal{C}_Y(V) > 0$, reverse the roles of X and Y in the construction, so that g maps $\mathcal{C}_Y(V)$ onto a subset of $\mathcal{C}_X(V)$.

The meeting of the chains
At the random time $T := \min\{t : D_t = \emptyset\}$ the colorings agree for all vertices. At that time both $\mathcal{C}_X(v)$ and $\mathcal{C}_Y(v)$ are empty for every v in $\mathcal{V}$. Subsequently, the algorithm merely modifies the color scheme without creating any new disagreements.

The details of Jerrum's argument appear in the following Problems.

Homework problems

Write f_t for $\#D_t$, the number of vertices where the colorings disagree after completion of step t. The main idea is to show that $\mathbb{E}f_{t+1} \le (1-\alpha)\mathbb{E}f_t$ where

$$\alpha := \frac{q - 2\Delta}{nq} > 0,$$

which will produce a rapidly decreasing bound on the total variation distance between π and the distribution of Y_t.

[1] (5 points) For each σ and τ in $\mathcal{S}$ show that $P(\sigma, \tau) = P(\tau, \sigma)$. Deduce that the uniform distribution π on $\mathcal{S}$ is the stationary distribution.

[2] (5 points) Explain why

$$\mathbb{TV}_t := \max_{B \subseteq \mathcal{S}} |\mathbb{P}\{Y_t \in B\} - \pi(B)| \le \mathbb{P}\{T > t\}$$

Hint: The argument is similar to the one used for the proof of the BLT.

[3] (10 points) Explain why $T = \min\{t : f_t = 0\}$ and

$$\mathbb{P}\{T > t\} \le \mathbb{P}\{f_t \ne 0\} \le \mathbb{E}f_t.$$

Hint: The random variable f_t takes only nonnegative integer values.

[4] (5 points) Explain why $|f_{t+1} - f_t| \le 1$ always, with $f_{t+1} \le f_t$ if $V_{t+1} \in D_t$ and $f_{t+1} \ge f_t$ if $V_{t+1} \in A_t$.

[5] (5 points) For each vertex v, define

$$d_t(v) := \begin{cases} \#(\mathcal{N}(v) \cap A_t) & \text{if } v \in D_t \\ \#(\mathcal{N}(v) \cap D_t) & \text{if } v \in A_t. \end{cases}$$

Equivalently, $d_t(v)$ is the number of edges (with v as one of the endpoints) that join a point in A_t to a point in D_t. Explain why the total number of edges that join a point in A_t to a point in D_t equals

$$m_t := \sum\nolimits_{v \in A_t} d_t(v) = \sum\nolimits_{v \in D_t} d_t(v).$$

[6] Consider the case where $V_{t+1} \in A_t$. Abbreviate V_{t+1} to V and C_{t+1} to C.

(i) (5 points) Explain why $f_{t+1} = f_t$ if $\mathcal{C}_X(V) = \mathcal{C}_Y(V) = \emptyset$.

Without loss of generality suppose $\#\mathcal{C}_X(V) \le \#\mathcal{C}_Y(V)$ and $\mathcal{C}_Y(V) \ne \emptyset$ for the rest of this problem. It is possible that $\mathcal{C}_X(V)$ is empty, which would make (ii) a bit easier.

(ii) (15 points) If $C \notin \mathcal{C}_Y(V)$, show that $f_{t+1} = f_t$. Hint: Consider separately the cases where C belongs to $(K_X(V) \cup K_Y(V))^c$ or $K_X(V) \cap K_Y(V)$ or $\mathcal{C}_X(V)$.

(iii) (5 points) Explain why f_{t+1} might equal $1 + f_t$ if $C \in \mathcal{C}_Y(V)$.

(iv) (5 points) Explain why $\#\mathcal{C}_Y(V) \le d_t(V)$ when $V \in A_t$.

(v) (10 points) Deduce that

$$\mathbb{P}\{f_{t+1} = 1 + f_t \mid \text{info}_t\} \le \mathbb{P}\{V \in A_t, C \in \mathcal{C}_Y(V) \mid \text{info}_t\} \le \frac{m_t}{nq}.$$

Explain your reasoning in detail.

[7] Consider the case where $V_{t+1} \in D_t$. Abbreviate V_{t+1} to V and C_{t+1} to C.

(i) (5 points) For each v in D_t explain why

$$\mathbb{P}(f_{t+1} = f_t - 1 \mid V = v, \text{info}_t) \ge \mathbb{P}(C \notin K_X(v) \cup K_Y(v) \mid V = v, \text{info}_t)$$

(ii) (5 points) Explain why $\#\left(K_X(v) \cup K_Y(v)\right) \le 2\Delta - d_t(v)$ for all v in D_t. Hint: What is the largest number of distinct colors that can be contributed by $\mathcal{N}(v) \cap D_t$?

(iii) Deduce that

$$\mathbb{P}(f_{t+1} = f_t - 1 \mid \text{info}_t) \ge \alpha f_t + \frac{m_t}{nq}$$

[8] Combine the last two results.

(i) (5 points) Show that

$$\begin{aligned}\mathbb{E}\left(f_{t+1} - f_t \mid \text{info}_t\right) &= \mathbb{P}\{f_{t+1} = 1 + f_t \mid \text{info}_t\} - \mathbb{P}(f_{t+1} = f_t - 1 \mid \text{info}_t)\\ &\leq -\alpha f_t\end{aligned}$$

(ii) (5 points) Deduce that $\mathbb{E} f_{t+1} \leq (1-\alpha)\mathbb{E} f_t$ for all t.

(iii) (5 points) Deduce that

$$\mathbb{E} f_t \leq (1-\alpha)^t \mathbb{E} f_0 \leq n(1-\alpha)^t.$$

Notice that the upper bound does not depend on μ.

(iv) (5 points) Conclude, via Problems [2] and [3] that

$$\mathbb{TV}_t \leq n(1-\alpha)^t$$

Remark. You could solve to find how large t must be in order to make $\mathbb{TV}_t \leq \epsilon$, for any given $\epsilon > 0$.

References

Jerrum, M. (1995). A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures and Algorithms 7*(2), 157–165.