Applications of Stochastics — Exercise sheet 5

Gábor Pete

http://www.math.bme.hu/~gabor

March 29, 2018

For a (possibly directed) graph, the adjacency matrix is $A_{u,v} = \mathbf{1}_{u \to v}$. The probability transition matrix for the corresponding Markov chain is $P_{u,v} = A_{u,v} / \sum_w A_{u,w}$. For an undirected graph on the vertex set $\{1, \dots, n\}$, we know from the Stochastic Processes course (and it is straightforward to verify) that P has a left eigenvector $\pi(i) = \deg(i)$, $1 \le i \le n$, with eigenvalue 1; i.e., it is a stationary measure. I mentioned in class, incorrectly, that the leading eigenvector of A is $\big(\sqrt{\deg(i)}\big)_i$. Here is the correct statement:

▷ **Exercise 1.**

(a) When P is the Markov transition matrix for any finite directed graph $G(V,E)$, show that $\|Pf\|_\infty \le \|f\|_\infty$ holds for any $f : V \longrightarrow \mathbb{R}$.

(b) Let A be the symmetric $n \times n$ adjacency matrix of an undirected finite graph on the vertex set $\{1, \dots, n\}$. Let D be the diagonal matrix formed by the degrees $\deg(i)$, and note that it is clear what $D^{-1/2}$ means. Show that all the eigenvalues of $B = D^{-1/2} A D^{-1/2}$ are real, are between 1 and -1, and that the vector $\big(\sqrt{\deg(i)}\big)_{1 \le i \le n}$ is an eigenvector for the eigenvalue 1. (Hint: use parts (a) and (c).)

(c) Observe that B from part (b) and the Markov transition matrix P are conjugate matrices, hence they have the same eigenvalues. Graph theorists prefer B to P because it is symmetric, and to A because it is normalized to have spectrum between 1 and -1.

We want to rank vertices of a directed graph according to importance. Here is a summary of what we did in class (clarifying why there is no need to talk about the leading eigenvalue for PageRank):

▷ **Exercise 2.**

(a) As a first idea, we used the iteration $\overline{x}_{t+1} := \overline{x}_t A$. Assume that A has a complete basis of eigenvectors $\overline{v}_i$, $i = 1, \dots, n$ (not at all the case in general), with a 1-dimensional eigenspace $\langle \overline{v}_1 \rangle$ corresponding to the eigenvalue λ_1 with the largest absolute value. Show that, for $\overline{x}_0 = \mathbf{1}$, there is a normalization c_t such that $\overline{x}_t / c_t$ converges to $\overline{v}_1$.

(b) In Google's **PageRank**, the iteration $\overline{x}_{t+1} := \alpha \overline{x}_t P + \mathbf{1}$ is used, with some $\alpha \in (0,1)$. Show that, for any starting vector $\overline{x}_0$, the sequence $\overline{x}_t$ converges to $\mathbf{1}\,(I - \alpha P)^{-1}$. (Hint: use the Banach fixed point theorem, with an appropriate notion of distance. See part (a) of Exercise 1.)

▷ **Exercise 3.** Consider the undirected graph on the vertex set $\{1,2,3,4\}$, where $1,2,3$ form a triangle, and 1 and 4 are also connected by an edge.

(a) Calculate the eigenvector importance from part (a) of the previous exercise.

(b) Calculate the PageRank scores from part (b) of the previous exercise, for several values of α.

You are welcome to use Mathematica or other software.

▷ **Exercise 4.** Recall that we defined the **clustering coefficient** of an undirected graph as

$$\mathsf{CC} := \frac{\#\text{ paths of length 2 with endpoints connected by an edge}}{\#\text{ paths of length 2}}\,.$$

With n vertices and $10n$ edges, find a graph with small CC, and another one with large CC.

As in class, let $\xi_1, \xi_2, \ldots$ be the i.i.d. lifetimes in a renewal process, with $\mathbf{E}\xi_i = \mu \in (0, \infty]$. Then $T_k := \sum_{i=1}^k \xi_i$ are the renewal times, and $N_t := \min\{k : T_k \geq t\}$. Also, $U(t) := \mathbf{E}N_t$, called the renewal function. The **Elementary Renewal Theorem** says that

$$\lim_{t\to\infty} \frac{U(t)}{t} = \frac{1}{\mu}\,. \tag{1}$$

The proof uses **Wald's identity**, a special case of an Optional Stopping Theorem, which we accepted without a proof: if $\mu < \infty$, and $\mathbf{E}N_t < \infty$ for any t, then

$$\mathbf{E}T_{N_t} = \mu\,\mathbf{E}N_t\,. \tag{2}$$

The ingredient $\mathbf{E}N_t < \infty$ was proved in class. Then, $T_{N_t} \geq t$ gives the lower bound $\lim_{t\to\infty} U(t)/t \geq 1/\mu$ (trivial when $\mu = \infty$). But I was puzzled why the books don't just use Fatou's lemma to get this. Well, the reason is that (2) is also needed to get the upper bound! Here it is:

$\triangleright$ **Exercise 5.** Consider the renewal process with $\overline{\xi_i} := \min\{\xi_i, K\}$ for any $K > 0$ fixed. Note that $\overline{T}_{\overline{N}_t} \leq t + K$, and get an upper bound for $\overline{U}(t)$. Then let $K \to \infty$ to get the upper bound for $U(t)$. Here, the key technical lemma that you should prove (then apply it to $a_K(t) := \overline{U}(t)/t$) is that if $a_K(t) \geq 0$, monotone decreasing in K for any fixed t, then

$$\limsup_{K\to\infty} \limsup_{t\to\infty} a_K(t) \geq \limsup_{t\to\infty} \limsup_{K\to\infty} a_K(t)\,.$$