Comparison of subdominant eigenvalues of some linear search schemes

Alan Pryde 17/07/2012

1. Linear Search Schemes

Suppose we have a collection of n items $B_1, B_2, \ldots, B_n$, such as files in a computer, ordered linearly from ”left” to ”right”. These items are accessed, independently in a statistical sense, with probabilities or weights $w_1, w_2, \ldots, w_n$. When an item is accessed the list is searched from left to right until the desired item is reached and then returned to the list according to various schemes.

This problem of dynamically organizing a linear list has been studied by probability theorists and computer scientists for many years. Two schemes that are frequently mentioned in the literature are the move-to-front and the transposition schemes.

In the **move-to-front scheme** the accessed item is returned to the front (left) of the list and all other items retain their relative positions.

In the **transposition scheme**, if the accessed item came from the front of the list then it is returned to the same position. Otherwise it is interchanged with the nearest item closer to the front of the list.

For each of these two schemes the successive configurations of the list of items forms a Markov chain whose state space is the symmetric group S_n of permutations of the numbers $1, 2, \ldots, n$.

The transition probability matrices for the move-to-front and transposition schemes, denoted Q and T respectively, are matrices indexed by the elements σ, τ of S_n.

2. Eigenvalues

Fact 1: If the weights are all positive, then Q and T are regular stochastic matrices and so the chains converge to stationary states. Their dominant (Perron) eigenvalues are $\mu_1(Q) = \mu_1(T) = w_1 + w_2 + \ldots + w_n = 1$.

Fact 2: The transposition chain is a reversible Markov chain $(\pi(\sigma)T(\sigma,\tau) = \pi(\tau)T(\tau,\sigma))$. Hence T has real eigenvalues.

Fact 3: The MTF matrix Q also has real eigenvalues. (See Theorem 1.)

The relative sizes of the subdominant eigenvalues $\mu_2(Q)$ and $\mu_2(T)$ are of interest because they determine the speed of transition to the stationary state.

$D(n)$ = the number of derangements of n elements

Recall that $\sum_{k=0}^{n} \binom{n}{k} D(n-k) = n!$

Theorem 1 ([1],[2]) For arbitrary complex weights the eigenvalues of Q are 0 with multiplicity $D(n)$ and the numbers $w_{i_1} + w_{i_2} + \ldots + w_{i_k}$ with multiplicity $D(n-k)$ where $1 \le i_1 < i_2 < \ldots < i_k \le n$, $1 \le k \le n$ and $k \ne n-1$.

Theorem 2 ([6]) For arbitrary non-negative weights, $\mu_2(T) \ge \mu_2(Q)$.

3. **Example** $n = 3$

Relative to reverse lexicographical order $(123, 213, 132, 312, 231, 321)$ the move-to-front t.p.m with weights a, b, c is given by

(σ, τ)	123	213	132	312	231	321
123	a	b	0	c	0	0
213	a	b	0	0	0	c
132	0	b	a	c	0	0
312	0	0	a	c	0	b
231	a	0	0	0	b	c
321	0	0	0	a	b	c

$$Q = \begin{bmatrix} a & b & 0 & c & 0 & 0 \\ a & b & 0 & 0 & 0 & c \\ 0 & b & a & c & 0 & 0 \\ 0 & 0 & a & c & b & 0 \\ a & 0 & 0 & 0 & b & c \\ 0 & 0 & a & 0 & b & c \end{bmatrix}$$

eigenvalues: $a + b + c, a, b, c, 0, 0$

$$T = \begin{bmatrix} a & b & c & 0 & 0 & 0 \\ a & b & 0 & 0 & c & 0 \\ b & 0 & a & c & 0 & 0 \\ 0 & 0 & a & c & 0 & b \\ 0 & a & 0 & 0 & b & c \\ 0 & 0 & 0 & a & b & c \end{bmatrix}$$

eigenvalues: $a + b + c, \ldots\ldots$

Question: Why are the eigenvalues of Q so simple and those of T so intractable?

4. **Some calculations**

Fact 4:

We write permutations in the form $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$ or $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$. Then

$$Q(\sigma, \tau) = \begin{cases} w_{\sigma(1)} & \text{if } \sigma = \tau \\ w_{\sigma(k)} & \text{if } \tau = (\sigma_k, \sigma_1, \ldots, \sigma_{k-1}, \sigma_{k+1}, \ldots, \sigma_n) \text{ for some } k > 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$T(\sigma,\tau) = \begin{cases} w_{\sigma(1)} & \text{if } \sigma = \tau \\ w_{\sigma(k)} & \text{if } \tau = (\sigma_1,\ldots,\sigma_{k-2},\sigma_k,\sigma_{k-1},\sigma_{k+1},\ldots,\sigma_n) \text{ for some } k > 1. \\ 0 & \text{otherwise} \end{cases}$$

Fact 5: Each row of both Q and T contains the weights $w_1, w_2, \ldots, w_n$ exactly once each, whereas the diagonals contain the weights exactly $(n-1)!$ times each.

Fact 2: The Markov chain for the transposition scheme is reversible.

Proof: Set $\pi(\sigma) = w_{\sigma(1)}^{n-1} w_{\sigma(2)}^{n-2} \ldots w_{\sigma(n-1)}^{1}$. Then $\pi(\sigma)T(\sigma,\tau) = \pi(\tau)T(\tau,\sigma)$ which is the defining condition for reversibility. In particular, summing over σ, we obtain $\pi T = \mu_1(T)\pi$ and so π is a stationary distribution for T in the case of probabilities $w_1, w_2, \ldots, w_n$ summing to 1.

Fact 6: If all weights are positive, T is similar to a symmetric matrix U.

Proof: Let R be the square diagonal matrix with $R(\sigma,\sigma) = \sqrt{\pi(\sigma)}$. Set $U = RTR^{-1}$. The reversibility condition becomes $T^t = R^2TR^{-2}$ and so $U^t = U$.

Fact 7: For non-negative weights, T has real eigenvalues.

Proof: A simple calculation shows that for positive weights

$$U(\sigma,\tau) = \begin{cases} w_{\sigma(1)} & \text{if } \sigma = \tau \\ \sqrt{w_{\sigma(k-1)}w_{\sigma(k)}} & \text{if } \tau = (\sigma_1,\ldots,\sigma_{k-2},\sigma_k,\sigma_{k-1},\sigma_{k+1},\ldots,\sigma_n) \text{ for some } k > 1. \\ 0 & \text{otherwise} \end{cases}$$

For the general case of non-negative weights R^{-1} may not exist so we define U by this last identity. By a simple continuity argument, T and U again have the same characteristic polynomial.

We will refer to U as the symmetrized form of T and sometimes write $U = U(w_1, w_2, \ldots, w_n)$ to denote its dependence on the weights.

For any matrix A with real eigenvalues, of size m by m say, we denote its eigenvalues by $\mu_1(A), \ldots, \mu_m(A)$ when arranged in decreasing order and by $\lambda_1(A), \ldots, \lambda_m(A)$ when the order is increasing.

5. Proof of theorem 2.

Theorem 2 For arbitrary non-negative weights, $\mu_2(T) \geq \mu_2(Q)$.

Proof Order the $n!$ row and column indices σ so that for the first $(n-1)!$ indices $\sigma(n) = n$, for the next $(n-1)!$ indices $\sigma(n) = n-1$ and so on. Then U has a block decomposition $U = [U_{ij}]$ for $1 \leq i,j \leq n$ whose diagonal blocks are of the form $U_{ii} = U(w_1, w_2, \ldots, \widehat{w}_{n+1-i}, \ldots, w_n)$. The symbol $\widehat{w}_j$ is used to denote that w_j is omitted. So $\mu_1(U_{ii}) = w_1 + w_2 + \ldots + \widehat{w}_{n+1-i} + \ldots + w_n$. For example, when $n = 3$:

$$U = \begin{bmatrix} a & \sqrt{ab} & \sqrt{bc} & 0 & 0 & 0 \\ \sqrt{ab} & b & 0 & 0 & \sqrt{ac} & 0 \\ \sqrt{bc} & 0 & a & \sqrt{ac} & 0 & 0 \\ 0 & 0 & \sqrt{ac} & c & 0 & \sqrt{ab} \\ 0 & \sqrt{ac} & 0 & 0 & b & \sqrt{bc} \\ 0 & 0 & 0 & \sqrt{ab} & \sqrt{bc} & c \end{bmatrix}.$$

To simplify notation we will assume that $w_1 \leq w_2 \leq \ldots \leq w_n$. As each U_{ii} is Hermitian, there are unitary matrices V_i such that each $V_i^* U_{ii} V_i$ is a diagonal matrix.

If $Z = diag(V_1, \ldots, V_n)$ then $Z^*Z = I$ and Z^*UZ is a block matrix whose diagonal blocks are diagonal matrices whose diagonal elements are the eigenvalues of the U_{ii}.

Now remove from Z the two columns corresponding to the Perron eigenvalues $\mu_1(U_{ii})$ for $i = n-1, n$ to obtain a non-square matrix $W = diag(W_1, \ldots, W_n)$. Then $W^*W = I_k$, the identity matrix of order $k = n! - 2$ and W^*UW is a block matrix whose diagonal blocks are diagonal matrices whose diagonal elements are the eigenvalues of the U_{ii} with the two Perron eigenvalues $\mu_1(U_{n-1,n-1})$ and $\mu_1(U_{nn})$ omitted.

So

$$\begin{aligned} trace(W^*UW) &= \sum_{i=1}^{n} trace(W_i^* U_{ii} W_i) \\ &= \sum_{i=1}^{n} trace(U_{ii}) - \mu_1(U_{n-1,n-1}) - \mu_1(U_{nn}) \\ &= trace(U) - (w_1 + \widehat{w}_2 + w_3 + \ldots + w_n) - (\widehat{w}_1 + w_2 + \ldots + w_n) \\ &= trace(U) - (w_3 + \ldots + w_n) - (w_1 + w_2 + \ldots + w_n) \\ &= trace(Q) - \mu_2(Q) - \mu_1(Q) \\ &= \sum_{i=1}^{n!-2} \lambda_i(Q). \end{aligned}$$

By the generalized Rayleigh-Ritz theorem (see Horn and Johnson 4.3.18) we have

$$\sum_{i=1}^{n!-2} \lambda_i(U) = \min\{trace(X^*UX) : X^*X = I_{n!-2}\}$$

and therefore

$$\sum_{i=1}^{n!-2} \lambda_i(T) = \sum_{i=1}^{n!-2} \lambda_i(U) \leq \sum_{i=1}^{n!-2} \lambda_i(Q).$$

Since T and Q have the same trace and the same Perron eigenvalue, we conclude that $\mu_2(T) \geq \mu_2(Q)$.

Using similar techniques, further information can be readily gained about the eigenvalues of T. For example:

Theorem 3 For non-negative weights, $\sum_{i=1}^{k} \lambda_i(T) \leq 0$ for $1 \leq k \leq n!/2$.

Proof Since the result is trivially true when $n = 2$, we may proceed by induction on n. Assume it is valid for lists of length $n-1$ for some $n > 2$. Take matrices U_{ii} as in the proof of Theorem 2. By the induction hypothesis and the generalized Rayleigh-Ritz theorem, for $1 \leq i \leq n$ and $1 \leq h \leq (n-1)!/2$ there are matrices W_{ih} of size $(n-1)! \times h$ with orthonormal columns such that $trace(W_{ih}^* U_{ii} W_{ih}) \leq 0$. Given $1 \leq k \leq n!/2$ choose integers $h_1, h_2, \ldots, h_m$ where $1 \leq m \leq n$, $1 \leq h_i \leq (n-1)!/2$ and $h_1 + h_2 + \ldots + h_m = k$. Let W be the $n! \times k$ block matrix whose diagonal blocks are $W_i = W_{ih_i}$ for $1 \leq i \leq m$ with zeros elsewhere. Then $W^*W = I_k$ and $trace(W^*UW) = \sum_{i=1}^{m} trace(W_i^* U_{ii} W_i) \leq 0$ so $\sum_{i=1}^{k} \lambda_i(T) \leq 0$.

Example 2 The situation is different if negative weights are permitted. For example, consider the case $n = 3$ and weights $-1, 2, 4$. The eigenvalues of Q are $5, 4, 2, -1, 0, 0$ and those of T are approximately $5.0, -3.429$, $3.128 \pm 1.283i, 1.086 \pm 1.643i$. So the eigenvalue with second largest modulus for Q is 4 and for T is -3.429.

6. A closer look at T and Q.

Let T_j and Q_j be the t.p.m s corresponding to weights $0, \ldots, 0, 1, 0, \ldots, 0$. Then

Fact 8

1. $T = w_1 T_1 + \ldots + w_n T_n$
2. $Q = w_1 Q_1 + \ldots + w_n Q_n$
3. $T_j^n = T_j^{n-1} = Q_j$
4. $Q_i\, Q_j(\sigma, \tau) = 1$ if τ can be obtained from σ by moving i to the front then j to the front and $Q_i\, Q_j(\sigma, \tau) = 0$ otherwise.
5. $Q_j^2 = Q_j$.
6. $Q_h Q_{j_1} Q_{j_2} \ldots Q_{j_k} Q_h = Q_{j_1} Q_{j_2} \ldots Q_{j_k} Q_h$
7. The semigroup generated by $Q_1, \ldots, Q_n$ consists of idempotents.
8. The algebra generated by $Q_1, \ldots, Q_n$ is triangularizable.
9. The eigenvalues of Q are of the form $w_1 \lambda_1 + \ldots + w_n \lambda_n$ where λ_j is an eigenvalue of Q_j namely 0 or 1.

References

[1] R.M. Phatarfod, *On the matrix occuring in a linear search problem*, J. Appl. Prob. 28 (1991), 336-346.

[2] R.M. Phatarfod, A.J. Pryde and D. Dyte, *On the move-to-front scheme with Markov dependent requests*, J. Appl. Prob. 34 (1997), 790-794.

[3] J.A. Fill, *An exact formula for the move-to-front rule for self-organising lists*, J. Theoretical Prob. 9 (1996), 113-160.

[4] A.J. Pryde and R.M. Phatarfod, *On some multi-request move-to-front heuristics*, J. Appl. Prob. 35 (1998), 911-918.

[5] P.R. Jelenković, A. Radovanović, *The persistent-access-caching algorithm*, Random Structures and Algorithms **33** (2008), 219-251.

[6] A.J. Pryde, *Comparison of subdominant eigenvalues of some linear search schemes*, Linear Algebra Appl. **431** (2009), 1439-1442