Statistics & Probability Letters 17 (1993) 163–168
North-Holland

26 May 1993

Sensitivity of finite Markov chains under perturbation

E. Seneta
School of Mathematics and Statistics, University of Sydney, NSW, Australia

Received September 1992
Revised November 1992

Abstract: Meyer (1992) has developed inequalities in terms of the non-unit eigenvalues λ_j, $j=2,\dots,n$, of a stochastic matrix P containing a single irreducible set of states, for the condition number $\max|a_{ij}^{\#}|$, where $A^{\#}=\{a_{ij}^{\#}\}$ is the group generalized inverse of $A=I-P$. In this note we derive, succinctly, analogous inequalities for the alternative condition number, the ergodicity coefficient $\tau_1(A^{\#})$, using the properties of ergodicity coefficients: $(\min|1-\lambda_j|)^{-1}\leqslant\tau_1(A^{\#})\leqslant\Sigma(1-\lambda_j)^{-1}$.

AMS Subject Classification: 15A51, 60J10, 65F15.

Keywords: Condition number; ergodicity coefficients; eigenvalues; group inverse sensitivity; perturbation of Markov chains; fundamental matrix; stochastic matrix; stationary distribution.

1. Introduction

If P is an $n\times n$ stochastic matrix containing a single irreducible set of states, so that there is a unique stationary distribution vector $\boldsymbol{\pi}^{\mathrm{T}}=\{\pi_i\}(\boldsymbol{\pi}^{\mathrm{T}}P=\boldsymbol{\pi}^{\mathrm{T}},\ \boldsymbol{\pi}^{\mathrm{T}}\mathbf{1}=1)$, and $\bar{P}$ is any other $n\times n$ stochastic matrix, with $\bar{\boldsymbol{\pi}}^{\mathrm{T}}=\{\bar{\pi}_i\}$ a stationary distribution vector corresponding to it, then direct matrix manipulation yields

$$(\bar{\boldsymbol{\pi}}^{\mathrm{T}}-\boldsymbol{\pi}^{\mathrm{T}})(I-P+\mathbf{1}\boldsymbol{\pi}^{\mathrm{T}})=\bar{\boldsymbol{\pi}}^{\mathrm{T}}E \tag{1}$$

where $E=\{e_{ij}\}=\bar{P}-P$. Then, denoting by $Z=\{z_{ij}\}=(I-P+\mathbf{1}\boldsymbol{\pi}^{\mathrm{T}})^{-1}$ the fundamental matrix (Kemeny and Snell, 1960) of the Markov chain governed by P, and by $A^{\#}=\{a_{ij}^{\#}\}=(I-P+\mathbf{1}\boldsymbol{\pi}^{\mathrm{T}})^{-1}-\mathbf{1}\boldsymbol{\pi}^{\mathrm{T}}=Z-\mathbf{1}\boldsymbol{\pi}^{\mathrm{T}}$ the group generalized inverse of $A=\{a_{ij}\}=I-P$, we obtain from (1), since $\bar{\boldsymbol{\pi}}^{\mathrm{T}}E\mathbf{1}=0$,

$$\bar{\boldsymbol{\pi}}^{\mathrm{T}}-\boldsymbol{\pi}^{\mathrm{T}}=\bar{\boldsymbol{\pi}}^{\mathrm{T}}EA^{\#}. \tag{2}$$

Thus

$$|\bar{\pi}_j-\pi_j|=\Big|\sum_i\sum_s\bar{\pi}_ie_{is}a_{sj}^{\#}\Big|\leqslant\Big(\sum_i\sum_s\bar{\pi}_i|e_{is}|\Big)\Big(\max_{k,j}|a_{kj}^{\#}|\Big)\leqslant\Big(\sum_i\bar{\pi}_i\Big)\Big(\max_r\sum_s|e_{rs}|\Big)\Big(\max_{k,j}|a_{kj}^{\#}|\Big)$$

so that (Funderlic and Meyer, 1986)

$$\max_j|\bar{\pi}_j-\pi_j|\leqslant\|E\|_1\max_{i,j}|a_{ij}^{\#}|. \tag{3}$$

Correspondence to: E. Seneta, School of Mathematics and Statistics, F07, University of Sydney, Sydney, NSW 2006, Australia.

(Equation (1) and equations close to (2) and (3) first occur in Schweitzer (1968).) On the other hand, from (2),

$$\| \bar{\pi}^{\mathrm{T}} - \pi^{\mathrm{T}} \|_1 = \| \bar{\pi}^{\mathrm{T}} E \|_1 \left(\| \bar{\pi}^{\mathrm{T}} E A^{\#} \|_1 / \| \bar{\pi}^{\mathrm{T}} E \|_1 \right) \leqslant \| \bar{\pi}^{\mathrm{T}} \|_1 \| E \|_1 \sup_{\| \delta^{\mathrm{T}} \|_1 = 1, \delta^{\mathrm{T}} \mathbf{1} = 0} \| \delta^{\mathrm{T}} A^{\#} \|_1$$

so that (Seneta, 1988, 1991)

$$\left(\| \bar{\pi}^{\mathrm{T}} - \pi^{\mathrm{T}} \|_1 / \| \pi^{\mathrm{T}} \|_1 \right) / \left(\| E \|_1 / \| P \|_1 \right) \leqslant \tau_1(A^{\#}) \tag{4}$$

since $\| \bar{\pi}^{\mathrm{T}} \|_1 = \| \pi^{\mathrm{T}} \|_1 = \| P \|_1 = 1$.

For any $n \times n$ matrix $B = \{B_{ij}\}$ with equal row sums b (i.e. $B\mathbf{1} = b\mathbf{1}$),

$$\tau_1(B) \overset{\text{def}}{=} \sup_{\| \delta^{\mathrm{T}} \| = 1, \delta^{\mathrm{T}} \mathbf{1} = 0} \| \delta^{\mathrm{T}} B \|_1 = \max_{i,j} \tfrac{1}{2} \sum_{s=1}^{n} | B_{is} - B_{js} | \tag{5}$$

(Seneta, 1984) and further, for any eigenvalue λ of B such that $\lambda \neq b$, it is true that

$$| \lambda | \leqslant \tau_1(B) \tag{6}$$

providing there is a left eigenvector v^{T} corresponding to b such that $v^{\mathrm{T}} \mathbf{1} = 1$ (Seneta, 1984). This property holds for the matrices $B = P$, Z, $I - P$, $A^{\#}$ with $v^{\mathrm{T}} = \pi^{\mathrm{T}}$ and $b = 1, 1, 0, 0$ respectively. These properties will be useful in the sequel.

On the basis of (3), Funderlic and Meyer (1986), Meyer (1992) choose

$$\kappa(A) = \max_{i,j} \left| a_{ij}^{\#} \right|$$

as a measure of relative sensitivity ('condition number') of π under perturbation of P to $\bar{P}$, while on the basis of (4) Seneta (1991) proposes $\tau_1(A^{\#}) \equiv \tau_1(Z)$ as condition number.

Recently, Meyer (1992) has addressed the question whether the closeness of the non-unit eigenvalues of P to unit provides complete information about the relative sensitivity of P. He has established that this is so by deriving the inequalities

$$\frac{1}{n \min_{2 \leqslant j \leqslant n} |1 - \lambda_j|} \leqslant \kappa(A) \leqslant \frac{2(n-1)\delta}{\prod_{j=2}^{n} (1 - \lambda_j)} \tag{7}$$

where $1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of P. Here

$$\delta = \max_{i,j,i \neq j} \prod_{k \neq i,j} a_{kk}$$

is the product of all but the two smallest diagonal entries of $A = I - P$. Note that since Z has the eigenvalues 1, $(1 - \lambda_j)^{-1}$, $j = 2, \ldots, n$,

$$\kappa(A) \leqslant 2(n-1)\delta \det Z. \tag{8}$$

We note that since the eigenvalues of $A^{\#}$ are 0, $(1 - \lambda_j)^{-1}$, $j = 2, \ldots, n$, and $A^{\#}\mathbf{1} = \mathbf{0}$, by (6),

$$\frac{1}{\min_{2 \leqslant j \leqslant n} |1 - \lambda_j|} \leqslant \tau_1(A^{\#}) \tag{9}$$

which is a slight sharpening of an inequality in Seneta (1991), produced at Meyer's instigation.

Thinking of $\tau_1(A^{\#})$ as condition number, we note that (9) has a sharper form than the left-hand side of (7). The question arises whether, using the properties (5) and (6) of the 'coefficient of ergodicity' $\tau_1(B)$, an upper bound resembling that in (7) can be obtained for $\tau_1(A^{\#})$ to yield the same *qualitative* conclustion as Meyer's. Such a question is now only of interest if this can be done simply, since the right-hand side of (7) required considerable technical ingenuity at some length. We are able to obtain quite quickly (a proof is provided in the next section) the upper bound

$$\tau_1(A^{\#}) \leqslant \sum_{j=2}^{n} \frac{1}{1-\lambda_j} = \operatorname{tr}(A^{\#}). \tag{10}$$

Since $\tau_1(A^{\#}) \geqslant 0$ (by (5)) it follows from (9) and (10) by the triangle inequality that

$$\frac{1}{\min_{2\leqslant j\leqslant n}|1-\lambda_j|} \leqslant \tau_1(A^{\#}) \leqslant \frac{n}{\min_{2\leqslant j\leqslant n}|1-\lambda_j|} \tag{11}$$

in parallel to (7). We shall see by examples that neither of the upper bounds (8) or (10) provides a particularly useful *quantitative* bound nor are they meant to. Bounds such as

$$\tau_1(A^{\#}) \leqslant (1-\tau_1(P))^{-1}$$

(Seneta, 1988, 1991) are much better quantitatively, and easily calculated from P by (5). (Indeed $0 \leqslant \tau_1(P) \leqslant 1$, and there are devices for coping with the situation where $\tau_1(P) = 1$.) On the other hand, the right-hand side of (10), in contrast to that of (7), shows that the situation noted by Meyer where no single λ_j is very close to 1, but enough are within range of 1 to force the right-hand side in (7) to be large does not affect the right-hand side of (10) in the same way, and confirms that such P are not badly conditioned.

2. Proof of (10)

Let $f_{ij}^{(k)}$, $k \geqslant 0$, be the probability of first passage from i to j in k steps ($f_{ij}^{(0)} = 0$), put $p_{ij}^{(0)} = 1$ if $i = j$, $= 0$ if $i \neq j$, and write for $|z| < 1$,

$$F_{ij}(z) = \sum_{k=0}^{\infty} f_{ij}^{(k)} z^k, \qquad P_{ij}(z) = \sum_{k=0}^{\infty} p_{ij}^{(k)} z^k.$$

Then it is clear that $F_{ij}(1) \leqslant 1$, and well-known (e.g. Seneta, 1981, Section 5.4) that $P_{ii}(z) = (1 - F_{ii}(z))^{-1}$, $P_{ij}(z) = F_{ij}(z)P_{jj}(z)$, $i \neq j$. Thus for $0 < \beta < 1$, for all i, j, $P_{ij}(\beta) \leqslant P_{jj}(\beta)$ whence

$$\sum_{k=0}^{\infty} \beta^k \left(p_{ij}^{(k)} - \pi_j\right) \leqslant \sum_{k=0}^{\infty} \beta^k \left(p_{jj}^{(k)} - \pi_j\right). \tag{12}$$

Now, according to Blackwell (1962), for our present structure of matrix P,

$$Z = I + \lim_{\beta \uparrow 1} \sum_{k=1}^{\infty} \beta^k (P^k - \mathbf{1}\boldsymbol{\pi}^{\mathrm{T}}),$$

so

$$A^{\#} = \lim_{\beta \uparrow 1} \sum_{k=0}^{\infty} \beta^k (P^k - \mathbf{1}\boldsymbol{\pi}^{\mathrm{T}}).$$

Using (12),

$$a_{ij}^{\#} \leqslant a_{jj}^{\#}. \tag{13}$$

Since $\frac{1}{2}|a-b| = \max(a, b) - \frac{1}{2}(a+b)$,

$$\tau_1(A^\#) = \max_{i,j} \tfrac{1}{2} \sum_{s=1}^{n} \left|a_{is}^\# - a_{js}^\#\right| = \max_{i,j} \sum_{s=1}^{n} \left\{\max\left(a_{is}^\#, a_{js}^\#\right) - \tfrac{1}{2}\left(a_{is}^\# + a_{js}^\#\right)\right\}$$

(repeating some lines from Seneta, 1981, p. 63). Since $A^\#$ has zero row sums,

$$\begin{aligned}\tau_1(A^\#) &= \max_{i,j} \sum_{s=1}^{n} \max\left(a_{is}^\#, a_{js}^\#\right)\\ &\leqslant \sum_{s=1}^{n} a_{ss}^\# \quad (\text{by } (13))\\ &= \operatorname{tr}(A^\#) \quad (\text{as required}). \qquad \square\end{aligned}$$

An alternative direct, but longer, proof of (13) is possible.

3. Examples

Example 1. For the 8×8 matrix P in Funderlic and Meyer (1986) (the Whittaker example), the largest non-unit eigenvalue is 0.911387, $\tau_1(P) = 0.912$, $\boldsymbol{\pi}^{\mathrm{T}} = (0.137, 0.049, 0.011, 0.014, 0.008, 0.050, 0.494, 0.238)$ and the matrix $A^\#$ (misprinted in Funderlic and Meyer) is

$$A^\# = \begin{bmatrix} 3.276 & 1.003 & -0.015 & -0.059 & 0.030 & -0.209 & -3.952 & -0.074\\ -0.329 & 2.943 & 0.005 & -0.035 & 0.275 & -0.121 & -3.084 & 0.344\\ -0.156 & -0.211 & 1.019 & 1.191 & 0.034 & -0.057 & -2.462 & 0.643\\ -0.299 & -0.262 & 0.007 & 2.989 & 0.253 & -0.110 & -2.976 & 0.396\\ -1.392 & -0.648 & -0.082 & -0.139 & 11.177 & -0.510 & -6.909 & -1.497\\ -0.360 & -0.283 & 0.002 & -0.038 & -0.128 & 3.714 & -3.196 & 0.290\\ -0.888 & -0.470 & -0.041 & -0.090 & -0.158 & -0.326 & 2.597 & -0.625\\ 0.167 & -0.097 & 0.045 & 0.014 & -0.098 & 0.061 & -1.297 & 1.204 \end{bmatrix}.$$

Thus

$$\kappa(A) = 11.18 \leqslant 2(n-1)\delta \det Z = 151.89,$$

$$\tau_1(A^\#) = 11.34 \leqslant \operatorname{tr}(A^\#) = 28.92,$$

$$\tau_1(A^\#) = 11.34 \leqslant (1 - \tau_1(P))^{-1} = 11.36,$$

clearly demonstrating the effectiveness of the simple bound $(1-\tau_1(P))^{-1}$. Note also that the left-hand side of (9) for this example is 11.285, so $\kappa(A)$ does not always bound the left-side of (9), and hence (7) cannot be sharpened to the same extent.

Example 2. The intention here is to construct a well-conditioned P for which none of the non-unit eigenvalues is close to unity, but their cumulative effect makes the right-hand side of (7) large, as envisaged by Meyer (1992),

$$P=\begin{bmatrix} 1-\varepsilon & \varepsilon/k & 0 & \varepsilon/k & 0 & \varepsilon/k & 0 & \cdots & \varepsilon/k & 0 \\ 1-a & 0 & a & 0 & 0 & 0 & 0 & & 0 & 0 \\ 1-a & a & 0 & 0 & 0 & 0 & 0 & & 0 & 0 \\ 1-a & 0 & 0 & 0 & a & 0 & 0 & & 0 & 0 \\ 1-a & 0 & 0 & a & 0 & 0 & 0 & & 0 & 0 \\ 1-a & 0 & 0 & 0 & 0 & 0 & a & & 0 & 0 \\ 1-a & 0 & 0 & 0 & 0 & a & 0 & & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-a & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & a \\ 1-a & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & a & 0 \end{bmatrix}.$$

P, aside from being bordered on the left and top, is block diagonal with k 2×2 blocks. Here $n=2k+1$ is taken large, a is taken nearly unity and $\varepsilon \ll 1$. We now show that the upper bound in (7) will diverge much faster than that in (10) or that below (11). Each of the k diagonal blocks has eigenvalues $\pm a$, so P has, apart from a unit eigenvalue, k eigenvalues almost a, and k eigenvalues almost $-a$ (letting $\varepsilon \to 0$). Also, $\delta = 1$, since $a_{11}=\varepsilon$, $a_{ii}=1$, $i=2,\dots,n$ $(A=I-P)$. Therefore,

$$2(n-1)\delta\prod_{j=2}^{n}\frac{1}{1-\lambda_j} \simeq 2(n-1)/(1-a^2)^{(n-1)/2}$$

while

$$\sum_{j=2}^{n}\frac{1}{1-\lambda_j} \simeq (n-1)/(1-a^2).$$

On the other hand, $\tau_1(P)=a$, so

$$\tau_1(A^{\#}) \leqslant (1-\tau_1(P))^{-1} = 1/(1-a).$$

Note that the case $\varepsilon=0$ is covered by our results; then $\boldsymbol{\pi}^{\mathrm{T}}=(1,0,0,\dots,0)$, the irreducible sub-set consisting of the first index only. When $\varepsilon>0$, the matrix is irreducible and aperiodic (primitive).

Acknowledgement

The author thanks the referee for his various suggestions.

References

Blackwell, D. (1962), Discrete dynamic programming, *Ann. Math. Statist.* **33** 719–726.

Funderlic, R.E. and C.D. Meyer (1986), Sensitivity of the stationary distribution vector for an ergodic Markov chain, *Linear Algebra Appl.* **76**, 1–17.

Kemeny, J.G. and L.J. Snell (1960), *Finite Markov Chains* (Van Nostrand, New Jersey).

Meyer, C.D. (1992), Sensitivity of Markov chains, submitted to: *SIAM J. Matrix Anal. Appl.*

Schweitzer, P.J. (1968), Perturbation theory and finite Markov chains, *J. Appl. Probab.* **5**, 401–413.

Seneta, E. (1979), Coefficients of ergodicity: structure and applications, *Adv. Appl. Probab.* **11**, 576–590.

Seneta, E. (1981), *Non-negative Matrices and Markov Chains* (Springer, New York, 2nd ed.).

Seneta, E. (1984), Explicit forms for ergodicity coefficients and spectrum localization, *Linear Algebra Appl.* **60**, 187–197.

Seneta, E. (1988), Perturbation of the stationary distribution measured by ergodicity coefficients, *Adv. Appl. Probab.* **20**, 228–230.

Seneta, E. (1991), Sensitivity analysis, ergodicity coefficients, and rank-one updates for finite Markov chains, in: W.J. Stewart, ed., *Numerical Solution of Markov Chains* (Dekker, New York) pp. 121–129.