SENSITIVITY OF MARKOV CHAINS

CARL D. MEYER

North Carolina State University

Raleigh, NC

7th Conference of the International

Linear Algebra Society

Madison, Wisconsin, June 3-6, 1998

Outline

- Background & History
- Norm-Based Analysis
- Ergodicity Coefficients
- Component-Wise Analysis
- First Passage Times Measures
- Comparisons & Conclusions

BACKGROUND

Finite, Homogeneous, Irreducible

- $\mathbf{P}_{n\times n}$ transition probability matrix
- $\boldsymbol{\pi}^T_{1\times n}$ stationary probability vector
- $\mathbf{A} = \mathbf{I} - \mathbf{P}$ irreducible M-matrix of rank $n-1$

$$\boldsymbol{\pi}^T\mathbf{A} = 0, \qquad \boldsymbol{\pi}^T\mathbf{e} = 1$$

Absolute Stochastic Perturbations

- $\widetilde{\mathbf{P}} = \mathbf{P} - \mathbf{E}$ stochastic & irreducible
- $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}$ irreducible M-matrix of rank $n-1$
- $\widetilde{\boldsymbol{\pi}}^T\widetilde{\mathbf{A}} = 0$ stationary probability vector

Question

- How different is $\widetilde{\boldsymbol{\pi}}^T$ from $\boldsymbol{\pi}^T$?

Straightforward?

Just An Ordinary Eigenvector Problem?

- $\boldsymbol{\pi}^T\mathbf{P} = \boldsymbol{\pi}^T, \qquad \boldsymbol{\pi}^T\mathbf{e} = 1$

  ✓ Use standard spectral perturbation theory

Just An Ordinary Linear System?

- Delete the last column of $\mathbf{A} = \mathbf{I} - \mathbf{P}$ and replace with 1's

  ✓ $\begin{cases} \text{Nonsingular linear system } \boldsymbol{\pi}^T\mathbf{N} = \mathbf{e}_n^T \\ \text{Use standard methods — } \kappa = \|\mathbf{N}\|\,\|\mathbf{N}^{-1}\| \end{cases}$

Why Not?

- Doesn't exploit structure — results are coarse
- Underlying mechanisms leading to sensitivity in Markov chains are not easily revealed

Basic Relationships

Combine $(\pi^T - \widetilde{\pi}^T)\mathbf{A} = \widetilde{\pi}^T\mathbf{E}$ **and** $(\pi - \widetilde{\pi})^T\mathbf{e} = 0$

- $$(\pi^T - \widetilde{\pi}^T)(\mathbf{A} + \mathbf{e}\pi^T) = \widetilde{\pi}^T\mathbf{E}$$

Kemeny's Fundamental Matrix

- $$\mathbf{Z} = (\mathbf{I} - \mathbf{P} + \mathbf{e}\pi^T)^{-1} = (\mathbf{A} + \mathbf{e}\pi^T)^{-1}$$

A Perturbation Formula

- $$(\pi^T - \widetilde{\pi}^T) = \widetilde{\pi}^T\mathbf{EZ}$$ (P. Schweitzer)

A Norm-Wise Bound

- $\|\pi^T - \widetilde{\pi}^T\| \leq \|\mathbf{E}\|\,\|\mathbf{Z}\|$ using $\|\star\|_1$ or $\|\star\|_\infty$

Group Inversion

For a matrix $\mathbf{T}$ having λ as a semi-simple eigenvalue:

- $0 \in \sigma(\mathbf{T} - \lambda\mathbf{I})$ is simi-simple $\Rightarrow$ $\mathbf{T} - \lambda\mathbf{I} = \mathbf{X}\begin{bmatrix} 0 & 0 \\ 0 & \mathbf{B} \end{bmatrix}\mathbf{X}^{-1}$
- $\mathbf{T} - \lambda\mathbf{I} \in$ multiplicative group $\mathcal{G}$
- Inverse of $\mathbf{T} - \lambda\mathbf{I} \in$ in $\mathcal{G}$ is $(\mathbf{T} - \lambda\mathbf{I})^{\#} = \mathbf{X}\begin{bmatrix} 0 & 0 \\ 0 & \mathbf{B}^{-1} \end{bmatrix}\mathbf{X}^{-1}$
- $(\mathbf{T} - \lambda\mathbf{I})^{\#}$ is Kato's reduced resolvent at $z = 0$

$$(z\mathbf{I} - (\mathbf{T} - \lambda\mathbf{I}))^{-1} = \frac{\mathbf{G}_{-1}}{z} + \mathbf{G}_0 + z\mathbf{G}_1 + z^2\mathbf{G}_2 + \cdots$$

$$(\mathbf{T} - \lambda\mathbf{I})^{\#} = \mathbf{G}_0 = \frac{1}{2\pi i}\oint_{C_{0'}} \frac{(z\mathbf{I} - (\mathbf{T} - \lambda\mathbf{I}))^{-1}}{z}$$

Derivatives Of Eigenvectors

For General T(z): Let $\mathbf{T}(z)\mathbf{x}(z) = \lambda(z)\mathbf{x}(z)$, $\mathbf{y}^*(z)\mathbf{T}(z) = \lambda(z)\mathbf{y}^*(z)$

$\lambda(z)$ simple, $\dot{\mathbf{T}}$, $\dot{\lambda}$, $\dot{\mathbf{x}}$, $\dot{\mathbf{y}}^*$ exist in a domain $\mathcal{D}$.

- If $\mathbf{x}^*\mathbf{x} = 1$ on $\mathcal{D}$, then

  ✓ $$\dot{\mathbf{x}} = \left[\mathbf{x}^*(\mathbf{T} - \lambda\mathbf{I})^{\#}\,\dot{\mathbf{T}}\,\mathbf{x}\right]\mathbf{x} + (\mathbf{T} - \lambda\mathbf{I})^{\#}\,\dot{\mathbf{T}}\,\mathbf{x}$$

- If $\mathbf{y}^*\mathbf{x} = 1$ on $\mathcal{D}$, then (G.W.Stewart & Meyer)

  ✓ $$\dot{\mathbf{x}} = -\left[\mathbf{x}^*\,\dot{\mathbf{y}}\right]\mathbf{x} - (\mathbf{T} - \lambda\mathbf{I})^{\#}\,\dot{\mathbf{T}}\,\mathbf{x}$$

For Markov Chains: If $\dot{\mathbf{P}}$ exists, then

✓ $$\dot{\boldsymbol{\pi}}^T = \boldsymbol{\pi}^T\,\dot{\mathbf{P}}\,\mathbf{A}^{\#}$$ (G.Golub & Meyer)

Another Approach

Another Fundamental Matrix

- $\mathbf{Z} = (\mathbf{A} + \mathbf{e}\pi^T)^{-1} = \mathbf{A}^\# + \mathbf{e}\pi^T$ suggests that we should

  ✓ use $\mathbf{A}^\#$ instead of $\mathbf{Z}$

Another Perturbation Formula

- $(\pi^T - \tilde{\pi}^T) = \tilde{\pi}^T \mathbf{E}\mathbf{A}^\#$ (Meyer)

Another Bound

- $\|\pi^T - \tilde{\pi}^T\| \leq \|\mathbf{E}\|\, \|\mathbf{A}^\#\|$ using $\|\star\|_1$ or $\|\star\|_\infty$

A Slightly Sharper Bound

(R.Funderlic & Meyer)

- $\|\pi^T - \tilde{\pi}^T\| \leq \|\mathbf{E}\| \max_{i,j} |\mathbf{A}^\#_{ij}|$ using $\|\star\|_\infty$

Improvements

Two Useful Facts

- $\mathbf{c}^T\mathbf{e} = 0 \Rightarrow |\mathbf{c}^T\mathbf{d}| \leq \|\mathbf{c}\|_1 \dfrac{d_{\max} - d_{\min}}{2} = \|\mathbf{c}\|_1 \max\limits_{i,k} \dfrac{|d_i - d_k|}{2}$
- Largest entry in each column of $\mathbf{A}^\#$ is on the diagonal

Apply To The Basic Relationship (with $\mathbf{c}^T = \widetilde{\boldsymbol{\pi}}^T\mathbf{E}$)

- $\|\boldsymbol{\pi}^T - \widetilde{\boldsymbol{\pi}}^T\|_\infty = \|\widetilde{\boldsymbol{\pi}}^T\mathbf{E}\mathbf{A}^\#\|_\infty = \max_j |(\widetilde{\boldsymbol{\pi}}^T\mathbf{E})\mathbf{A}^\#_{*j}|$
- $|(\widetilde{\boldsymbol{\pi}}^T\mathbf{E})\mathbf{A}^\#_{*j}| \leq \|\widetilde{\boldsymbol{\pi}}^T\mathbf{E}\|_1 \dfrac{\max_i \mathbf{A}^\#_{ij} - \min_i \mathbf{A}^\#_{ij}}{2}$

✓ $$\|\boldsymbol{\pi}^T - \widetilde{\boldsymbol{\pi}}^T\|_\infty \leq \|\mathbf{E}\|_\infty \max_j \frac{\mathbf{A}^\#_{jj} - \min_i \mathbf{A}^\#_{ij}}{2}$$

(Haviv & Van der Heyden)

✓ $$\|\boldsymbol{\pi}^T - \widetilde{\boldsymbol{\pi}}^T\|_\infty \leq \|\mathbf{E}\|_\infty \max_j \frac{\max_{i,k} |\mathbf{A}^\#_{ij} - \mathbf{A}^\#_{kj}|}{2}$$

(Kirkland, Neumann, Shader)

Ergodicity Coefficients

General Definition

- $$\tau_p(\mathbf{P}) \equiv \max_{\substack{\|\mathbf{y}^T\|_p=1 \\ \mathbf{y}^T\mathbf{e}=0}} \|\mathbf{y}^T\mathbf{P}\|_p$$

For Matrices With Constant Row Sums $(\mathbf{Be} = \beta\mathbf{e}, \;\; \mathbf{Ce} = \gamma\mathbf{e})$

- $\tau_1(\mathbf{B}) = \frac{1}{2}\max_{i,j}\|\mathbf{B}_{i*} - \mathbf{B}_{j*}\|_1 = \beta - \min_{i,j}\sum_{k=1}^{n}\min\{b_{ik}, b_{jk}\}$
- $\tau_1(\mathbf{BC}) \leq \tau_1(\mathbf{B})\tau_1(\mathbf{C})$
- $|\beta_2| \leq \tau_1(\mathbf{B}) \leq \|\mathbf{B}\|_1$ for $\beta_2 \neq \beta$ (Bauer, Deutch, Stoer)

For Stochastic Matrices

- $0 \leq \tau_1(\mathbf{P}) \leq 1$ $(= 1$ iff $\exists$ pair of $\perp$ rows)
- $\tau_1(\mathbf{P}) = 0$ if and only if $\mathbf{P} = \mathbf{e}\boldsymbol{\pi}^T$
- $\|\mathbf{RP}\|_\infty \leq \|\mathbf{R}\|_\infty\,\tau_1(\mathbf{P})$ whenever $\mathbf{Re} = 0$

Ergodicity Bounds

If $\mathbf{Re} = 0$ **and** $\tau_1(\mathbf{P}) \neq 1$

- $$\|\mathbf{RA}^{\#}\|_\infty \leq \frac{\|\mathbf{R}\|_\infty}{1 - \tau_1(\mathbf{P})}$$

Apply To Basic Relationship: $(\boldsymbol{\pi}^T - \widetilde{\boldsymbol{\pi}}^T) = \widetilde{\boldsymbol{\pi}}^T \mathbf{EA}^{\#}$

- $$\|\boldsymbol{\pi}^T - \widetilde{\boldsymbol{\pi}}^T\|_\infty \leq \frac{\|\mathbf{E}\|_\infty}{1 - \tau_1(\mathbf{P})}$$ (E.Seneta)

Improvements

- $$\tau_1(\mathbf{A}^{\#}) \leq \frac{1}{1 - \tau_1(\mathbf{P})} \quad \text{and} \quad \tau_1(\mathbf{A}^{\#}) \leq trace(\mathbf{A}^{\#})$$
- $$\|\boldsymbol{\pi}^T - \widetilde{\boldsymbol{\pi}}^T\|_\infty \leq \|\mathbf{E}\|_\infty \tau_1(\mathbf{A}^{\#}) \leq \|\mathbf{E}\|_\infty \, trace(\mathbf{A}^{\#})$$ (E.Seneta)

Sensitivity & Eigenvalues

For General Matrices

- Poorly separated eigenvalues $\Rightarrow$ sensitive eigenvector
- Well separated eigenvalues $\not\Rightarrow$ insensitive eigenvector

$$\mathbf{T} = \begin{bmatrix} 1 & -1 & -1 & \cdots & -1 \\ 0 & 2 & -2 & \cdots & -2 \\ 0 & 0 & 3 & \cdots & -3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \color{red}{0} & 0 & 0 & \cdots & n \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

It's Different For Stochastic Matrices

- $$\|\boldsymbol{\pi}^T - \widetilde{\boldsymbol{\pi}}^T\|_\infty \le \|\mathbf{E}\|_\infty \frac{2(n-1)\max_{i,j}\prod_{k\neq i,j} a_{kk}}{(1-\lambda_2)(1-\lambda_3)\cdots(1-\lambda_n)} \qquad \text{(Meyer)}$$
- $$\|\boldsymbol{\pi}^T - \widetilde{\boldsymbol{\pi}}^T\|_\infty \le \|\mathbf{E}\|_\infty \frac{n}{\min_{i\neq 1}|1-\lambda_i|} \qquad \text{(E. Seneta)}$$

Component-Wise Bounds

Absolute Error Bounds

- $$|\pi_j - \tilde{\pi}_j| \leq \|\mathbf{E}\|_\infty \max_i |\mathbf{A}^{\#}_{ij}|$$

(Funderlic&Meyer)

- $$|\pi_j - \tilde{\pi}_j| \leq \|\mathbf{E}\|_\infty \frac{\mathbf{A}^{\#}_{jj} - \min_i \mathbf{A}^{\#}_{ij}}{2} = \|\mathbf{E}\|_\infty \frac{\max_{i,k} |\mathbf{A}^{\#}_{ij} - \mathbf{A}^{\#}_{kj}|}{2}$$

(Haviv, Van der Heyden, Kirkland, Neumann, Shader)

Relative Error

- $$\frac{\pi_j - \tilde{\pi}_j}{\pi_j} = \tilde{\boldsymbol{\pi}}^T \mathbf{E}^{(j)} \mathbf{A}_j^{-1} \mathbf{e} \qquad \begin{cases} \mathbf{E}^{(j)} = \mathbf{E} \text{ with col } j \text{ deleted} \\ \mathbf{A}_j = j^{th} \text{ principal submatrix} \end{cases}$$

(Ipsen & Meyer)

Relative vs. Absolute Stability

Relative Error

- $$\left|\frac{\pi_j - \tilde{\pi}_j}{\pi_j}\right| \leq \rho_j \, \|\mathbf{E}\| \qquad \text{where} \qquad \rho_j = \frac{\|\mathbf{A}_j^{-1}\|_\infty}{\mathbf{2}}$$
- Equality is possible for all **P**

(Kirkland, Neumann, Shader, Ipsen, Meyer)

Absolute Error

- $$|\pi_j - \tilde{\pi}_j| \leq \left(\min_i \rho_i\right) \|\mathbf{E}\|_\infty$$

Conclusion

✓ If any one π_j is relatively well conditioned

then all π_j's are absolutely insensitive.

(Ipsen & Meyer)

Structured Perturbations

Relative Entry-Wise Perturbations

- $$\left|\frac{a_{ij} - \tilde{a}_{ij}}{a_{ij}}\right| \le \epsilon \quad \Longrightarrow \quad \left|\frac{\pi_j - \tilde{\pi}_j}{\pi_j}\right| \le \mathbf{2}n\epsilon + O(\epsilon^{\mathbf{2}})$$

(O'Cinneide)

✓ Small relative entry-wise perturbations produce only small relative errors in the π_j's

Different Interpretations Of Sensitivity

- $$\mathbf{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{bmatrix} \Rightarrow \mathbf{A} = \begin{bmatrix} \alpha & -\alpha \\ -\alpha & \alpha \end{bmatrix} \qquad 0 < \alpha << 1$$

✓ π_j's sensitive to small perturbations relative to 1

✓ π_j's not sensitive to relative entry-wise perturbations

Probabilistic Intrepretations

Mean First Passage Times

- Let M_{ij} = E[# steps to hit j for first time / start in i]

- $\mathbf{A}^{\#}_{jj} - \min_i \mathbf{A}^{\#}_{ij} = \max_i \frac{M_{ij}}{M_{jj}}$

✓ $$\begin{cases} |\pi_j - \tilde{\pi}_j| \leq \frac{\|\mathbf{E}\|_\infty}{\mathbf{2}} \max_i \frac{M_{ij}}{M_{jj}} \\ \left|\frac{\pi_j - \tilde{\pi}_j}{\pi_j}\right| \leq \frac{\|\mathbf{E}\|_\infty}{\mathbf{2}} \max_i M_{ij} \end{cases}$$

✓ π_j is insensitive if state j is “close” to all other states

✓ Suspect trouble if a state is “far removed” from others

✓ Chain is insensitive when $\exists$ strongly accessible state

Conclusions

- Markov chain perturbation analysis is more specialized than the theory for eigensystems and linear equations
- Inherent stochastic structure produces stronger (and surprising) conclusions about sensitivity
- Several different ways to construct perturbation bounds
- Sensitivity results depend on whether perturbations are measured absolutely or relatively
- All absolute perturbation bounds convey the same qualitative information — pick the one that best suits your application (e.g., numerical, modeling, etc.)