Shrinkage Inverse Wishart

김성민

서울대학교
통계학과, 베이즈통계 연구실

2023. 06. 15

목차

목차

참고문헌

- Gupta, A. K., and Nagar, D. K. *Matrix variate distributions*, vol. 104.
  CRC Press, 1999 [3]
- Chikuse, Y. *Statistics on special manifolds*, vol. 1.
  Springer [2]
- Hoff, P. D. Simulation of the matrix bingham–von mises–fisher distribution, with applications to multivariate and relational data.
  Journal of Computational and Graphical Statistics 18, 2 (2009), 438–456 [4]
- BERGER, J. O., SUN, D., and SONG, C. Bayesian analysis of the covariance matrix of a multivariate normal distribution with a new class of priors.
  The Annals of Statistics 48, 4 (2020), 2381–2403 [1]

목차

사전분포1

- Constant

$$\pi^C(\Sigma) \propto 1$$

- Jeffreys prior

$$\pi^J(\Sigma) \propto |\Sigma|^{-(p+1)/2}$$

- reference prior

$$\pi^R(\Sigma) \propto \frac{1}{|\Sigma| \prod_{i<j}(\lambda_i - \lambda_j)}$$

- modified reference prior

$$\pi^R(\Sigma) \propto \frac{1}{|\Sigma|^{1-1/(2p)} \prod_{i<j}(\lambda_i - \lambda_j)}$$

사전분포2

- Inverse Wishart (IW) prior

$$\pi^{IW}(\Sigma|a, H) \propto \frac{1}{|\Sigma|^a} etr(-\frac{1}{2}\Sigma^{-1}H)$$

- Shrinkage Inverse Wishart (SIW) prior

$$\pi^{SIW}(\Sigma|a, H) \propto \frac{1}{|\Sigma|^a \prod_{i<j}(\lambda_i - \lambda_j)} etr(-\frac{1}{2}\Sigma^{-1}H)$$

사전분포 형태

$$\pi(\Sigma|a,b,H) \propto \frac{1}{|\Sigma|^a[\prod_{i<j}(\lambda_i - \lambda_j)]^b} etr(-\frac{1}{2}\Sigma^{-1}H)$$

- $a = b = 0, H = 0$: Constant
- $a = (p+1)/2, b = 0$: Jeffreys prior
- $a = b = 1, H = 0$: reference prior
- $b = 0$: Inverse Wishart
- $b = 1$: Shrinkage Inverse Wishart

목차

변수 변환

$$\pi(\Sigma|a,b,H) \propto \frac{1}{|\Sigma|^a[\prod_{i<j}(\lambda_i-\lambda_j)]^b} etr(-\frac{1}{2}\Sigma^{-1}H)$$

- $\Sigma = \Gamma\Lambda\Gamma^T, \quad \Lambda = diag(\lambda_1,\ldots,\lambda_p)$

$$\left|\frac{\partial\Sigma}{\partial(\Gamma,\Lambda)}\right| = \prod_{i<j}(\lambda_i-\lambda_j)$$

Shrinkage

- 앞의 새로운 클래스의 사전분포는 다음과 같이 쓸 수 있다.

$$\pi(\Gamma, \Lambda | a, b, H) \propto \frac{etr(-\frac{1}{2}\Gamma\Lambda^{-1}\Gamma^T H)}{|\Lambda|^a [\prod_{i<j}(\lambda_i - \lambda_j)]^{b-1}} I_{\{\lambda_1 > \cdots > \lambda_p\}}$$

- $b = 0$ 인 경우를 생각해보면 고유값의 거의 같을 경우 사전분포가 거의 0이 된다. 즉, 해당 사전분포는 고유값이 퍼지게 만든다.
- $b = 1$ 인 경우에는 고유값이 얼마나 퍼져야 하는지에 대한 텀이 존재하지 않는다. $b = 0$ 인 경우에 비해서 덜 퍼지므로 "Shrinkage" 라고 표현한다.

정리3 [1]

SIW(a,H) 사전분포를 고려하자. $p = rank(H)$ 이고 $H = Z\Delta Z^T$로 스펙트럼 분해된다면 정수 $q \geq -1$에 대해서 다음이 성립한다.

$$\mathbb{E}[\Sigma^q] = Zdiag(\phi_{q,1}, \ldots, \phi_{q,p})Z^T$$

$$\phi_{q,1}(a, \Delta) = \frac{k\Gamma(a-q-1)}{2^q\Gamma(a-1)} \frac{\int t_{i1}^2 ||\bar{t}_1||^{2q} \prod_{j=1}^p ||\bar{t}_j||^{-2(a-1)} dT}{\int t_{i1}^2 \prod_{j=1}^p ||\bar{t}_j||^{-2(a-1)} dT}.$$

$T = (t_{ij})$는 직교정규 행렬이고, $||\bar{t}_j||^2 = \sum_{h=1}^p \delta_h t_{hj}^2$ 이다. 여기서 δ_h는 Δ 의 대각 성분이다.

목차

사후분포

$$y_i \overset{iid}{\sim} N(0, \Sigma) \; i = 1, \ldots, n$$
$$\Sigma \sim SIW(a, H).$$

$$\pi^{SIW}(\Sigma | a, H) \propto \frac{1}{|\Sigma|^a \prod_{i<j}(\lambda_i - \lambda_j)} etr(-\frac{1}{2}\Sigma^{-1}H)$$

- $H_0 = H + nS$ 라고 하면 사후분포는 $SIW(a + n/2, H_0)$ 이다.

Metropolis-Hasting Algorithm

- 1단계 : $\Sigma^* \sim IW(\frac{n+p+1}{2}, H_0)$
- 2단계 : λ_i^*, λ_i 가 각각 Σ^*와 Σ_t 의 고유값이라고 하자. 합격확률 α는 다음과 같다.

$$\alpha = \min \left\{ 1, \prod_{i<j} \frac{\lambda_i^t - \lambda_j^t}{\lambda_i^* - \lambda_j^*} \cdot \prod_{i=1}^{p} \left(\frac{\lambda_i^*}{\lambda_i^t} \right)^{\frac{p+1-2a}{2}} \right\}$$

- 3단계 : $\Sigma_{t+1} = \begin{cases} \Sigma^* & \text{with probability } \alpha, \\ \Sigma_t & \text{otherwise.} \end{cases}$

- 차원이 클 경우에는 제대로 수렴하지 않음.

방법2

Gibbs sampling

$$\pi(\Lambda, \Gamma | H_0) \propto \frac{etr(-\frac{1}{2}\Lambda^{-1}\Gamma^T H_0 \Gamma)}{\prod_{i=1}^{p} {\lambda_i}^r}, \quad r = a + n/2, \; H_0 = H + nS$$

- 1단계 : $\pi(\Lambda|\Gamma, H_0) \propto \prod_{i=1}^{p} \frac{1}{\lambda_i^r} e^{-c_i/\lambda_i}$,
  c_i는 $\Gamma^T H_0 \Gamma/2$ 의 (i, i) 성분이다.
- 2단계 : $\pi(\Gamma|\Lambda, H_0) \propto etr(-\frac{1}{2}\Lambda^{-1}\Gamma^T H_0 \Gamma)$

- 2단계에서 Γ에 대한 조건부 분포에서 샘플링하기 위해서, Γ의 임의의 두 행씩 업데이트하는 방법을 이용할 것이다.

Γ 샘플링

- $\pi(\Gamma|\Lambda, H_0) \propto etr(-\frac{1}{2}H_0\Gamma\Lambda^{-1}\Gamma^T)$ 로 쓸 수 있다.
- $H_0 = diag(h_1, \ldots, h_p)$ 라고 가정하고,
  $H_1 = diag(h_1, h_2)$, $H_2 = diag(h_3, \ldots, h_p)$ 라 하자.
- Γ 의 첫 두 행에 대한 조건부 분포는 다음과 같다.

$$
\begin{aligned}
&\pi(\Gamma_{12}|\Gamma_{-12}, \Lambda, H_0) \\
&\quad \propto etr\left\{ -\frac{1}{2}\begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix}\begin{pmatrix} \Gamma_{12} \\ \Gamma_{-12} \end{pmatrix}\Lambda^{-1}(\Gamma_{12}^T, \Gamma_{-12}^T)\right\} \\
&\quad \propto etr\left\{ -\frac{1}{2}H_1\Gamma_{12}\Lambda^{-1}\Gamma_{12}^T\right\}
\end{aligned}
$$

Γ 샘플링

- Γ 의 첫 두 행 업데이트를 다음과 같이 표현할 수 있다.

$$\Gamma^* = diag(X, I_{p-2})(\Gamma_{12}^T, \Gamma_{-12}^T)^T.$$

여기서 Γ_{12}는 Γ의 처음 두 행을 의미하고, Γ_{-12}는 나머지 $(p-2)$ 행을 나타낸다.

- $X = D_\epsilon X_\theta \equiv \begin{pmatrix} \epsilon_1 & 0 \\ 0 & \epsilon_2 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. 여기서 $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2}]$ 이고 $\epsilon_i = \pm 1$ 이다.

Γ 샘플링

- $\Gamma_{12} = D_\epsilon X_\theta T_{12}$.
  여기서 T_{12}는 $(T_{12}^T, \Gamma_{-12}^T)^T \in \mathbb{O}(p)$ 를 만족하는 행렬
- $\Gamma_{12} \to \epsilon, \theta$ 는 1-1 대응이다.

$$\begin{aligned}
&\pi(\theta, \epsilon | \Gamma_{-12}, \Lambda, H_0) \\
&\quad \propto \pi(\Gamma_{12} | \Gamma_{-12}, \Lambda, H_0) J(\Gamma_{12} \to \epsilon, \theta) \\
&\quad \propto etr\left\{ -\frac{1}{2} H_1 X_\theta T_{12} \Lambda^{-1} T_{12}^T X_\theta^T \right\} J(\Gamma_{12} \to \epsilon, \theta)
\end{aligned}$$

Jacobian 계산 1

- $\Gamma_{12} = D_\epsilon X_\theta T_{12}, \quad (T_{12}^T, \Gamma_{-12}^T)^T \in \mathbb{O}(p)$
- $D_\epsilon X_\theta \equiv \begin{pmatrix} \epsilon_1 & 0 \\ 0 & \epsilon_2 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \theta \in (-\frac{\pi}{2}, \frac{\pi}{2}], \ \epsilon_i = \pm 1.$

$$\begin{aligned} J(\Gamma_{12} \to \epsilon, \theta) &= J(\Gamma_{12} \to X_\theta, D_\epsilon) J(X_\theta \to \theta) J(D_\epsilon \to \epsilon) \\ &= J(\Gamma_{12} \to X_\theta, D_\epsilon) \end{aligned}$$

Jacobian 계산 2

- CHIKUSE, Y. *Statistics on special manifolds*, vol. 1. Springer [2] 의 정리 3.3.1 적용.
- $\Gamma^T = ((D_\epsilon X_\theta T_{12})^T, \Gamma_{-12}^T), \quad (T_{12}^T, \Gamma_{-12}^T)^T \in \mathbb{O}(p)$

$$
\begin{aligned}
[d\Gamma] &= [d\Gamma_{-12}][d(D_\epsilon X_\theta)] \\
&= [d\Gamma_{-12}][d(D_\epsilon X_\theta)] \\
&= [d\Gamma_{-12}]J(D_\epsilon \to \epsilon)(d\epsilon)J(X_\theta \to \theta)(d\theta) \\
&= [d\Gamma_{-12}](d\epsilon)(d\theta)
\end{aligned}
$$

$$
\therefore \quad \pi(\theta, \epsilon | \Gamma_{-12}, \Lambda, H_0) \propto etr\left\{ -\frac{1}{2} H_1 X_\theta T_{12} \Lambda^{-1} T_{12}^T X_\theta^T \right\}
$$

$$
\Gamma_{12}^{new} = H_\theta D_\epsilon T_{12}.
$$

Γ 샘플링

- $\Gamma_{12}\Lambda^{-1}\Gamma_{12}^{T} = \begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix}$ 여기서 $w \in (-\pi/2, \pi/2]$이고 $s_1 > s_2$ 이다.
- 간단한 계산을 통해서 다음을 얻을 수 있다. $(c_0 = -\frac{1}{2}(s_1 - s_2)(h_1 - h_2) \leq 0)$

$$\pi(\theta|\Gamma_{-12}, \Lambda, H_0) \propto \exp\left(c_0 \cos^2(\theta + \omega)\right)$$

- $\alpha = \cos^2(\theta + \omega)$ 라고 하면 다음을 얻을 수 있다.

$$\pi(\alpha|\Gamma_{-12}, \Lambda, H_0) \propto e^{c_0\alpha}\alpha^{-\frac{1}{2}}(1-\alpha)^{-\frac{1}{2}}$$

- 위의 조건부 분포는 Rejection sampling 을 통해 샘플링 가능하다.

목차

세팅

$$y_i \overset{iid}{\sim} N(0, \Sigma_p)\ i = 1, \ldots, n, \quad \Sigma_p = diag(p, \ldots, 1)$$
$$\Sigma_p \sim \pi(\Sigma_p).$$

- $p = 20,\ n = 20, 100, 200, 400, 1000, 2000$ 인 경우에 대해서 각각 200 번씩 반복.
- 사전분포로는 $SIW(4, 4I_p)$ 와 $IW(\alpha, \beta I_p)$ 로 설정을 하였다.
- α와 β 는 1,2차 적률이 동일하도록 설정하였다.

$$\alpha = \frac{3}{2} + \frac{5}{4}p + \frac{1}{4}\sqrt{(2+p)^2 + 16}$$
$$\beta = 2(\alpha - p - 1).$$

관측값

- 가장 큰 고유값 : $\lambda_1 = 20$
- 고유값의 합 : $tr(\Sigma) = 210$
- 사후분포에서 샘플링한 공분산의 (λ_1에 대응되는)고유벡터와 실제 공분산 고유벡터의 내적 : Angle(est)
- 표본 공분산(S)의 (λ_1에 대응되는)고유벡터와 실제 공분산 고유벡터의 내적 : Angle(sample)
- $R_1(\Sigma, \hat{\Sigma}) = \mathbb{E}[L_1(\Sigma, \hat{\Sigma})]$ - Entropy loss

$$L_1(\Sigma, \hat{\Sigma}) = tr(\hat{\Sigma}\Sigma^{-1}) - \log|\hat{\Sigma}\Sigma|^{-1} - p$$

- $R_2(\Sigma, \hat{\Sigma}) = \mathbb{E}[L_2(\Sigma, \hat{\Sigma})]$ - Quadratic loss

$$L_2(\Sigma, \hat{\Sigma}) = tr(\hat{\Sigma}\Sigma^{-1} - I_p)^2$$

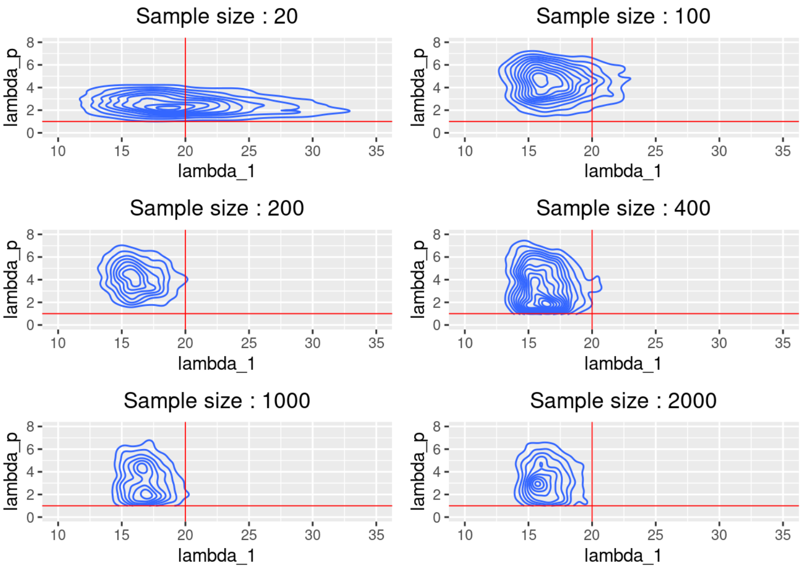
결과1 - SIW

n/p	lambda_1	tr(Sigma)	Angle(est)	Angle(sample)	R1	R2
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	21.00500	176.8902	0.2097803	0.3593252	63.83543	9.074997
5	17.70882	202.7790	0.2410482	0.4448114	69.94325	9.375340
10	17.03081	206.7965	0.2645120	0.5266634	57.40785	8.395700
20	16.71064	208.5116	0.2907960	0.6247529	46.88357	7.455083
50	16.37698	209.3859	0.3182361	0.7271319	41.90823	6.954476
100	16.17075	209.6797	0.3331834	0.8167924	43.98498	7.128943

결과1 - IW

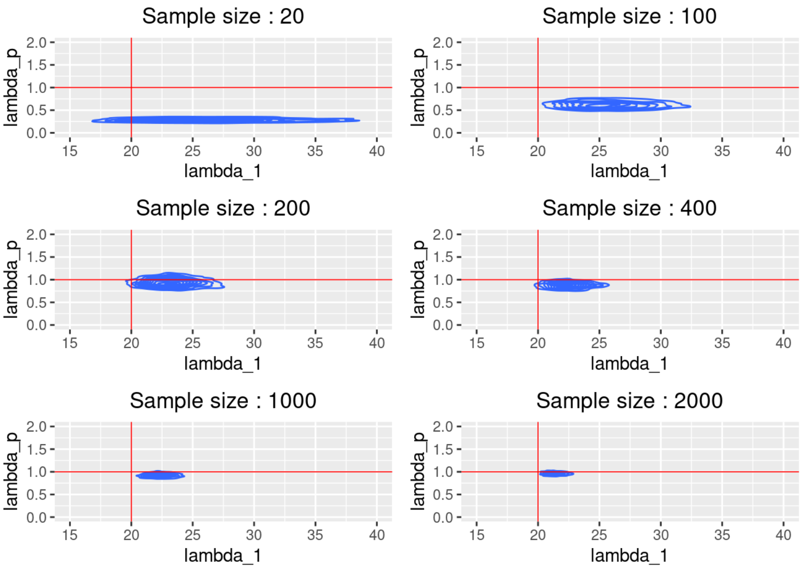
n/p <dbl>	lambda_1 <dbl>	tr(Sigma) <dbl>	Angle(est) <dbl>	Angle(sample) <dbl>	R1 <dbl>	R2 <dbl>
1	26.45256	109.2653	0.3123071	0.3593252	8.7685023	11.9810721
5	26.56511	175.5662	0.3783114	0.4448114	3.2575646	2.3150081
10	24.69294	191.4928	0.4446261	0.5266634	1.8255976	1.1043312
20	23.16580	200.4116	0.5294615	0.6247529	0.9884751	0.5467132
50	21.52570	206.0189	0.6256277	0.7271319	0.4135898	0.2157225
100	20.84336	207.9712	0.7310734	0.8167924	0.2083001	0.1063894

고유값 - SIW



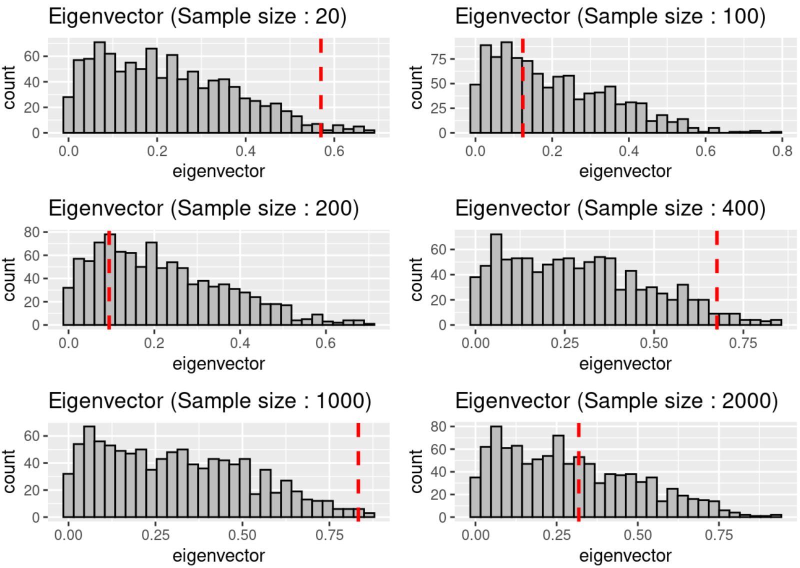
Sample size : 20
Sample size : 100
Sample size : 200
Sample size : 400
Sample size : 1000
Sample size : 2000
lambda_p
lambda_1
0
2
4
6
8
10
15
20
25
30
35

고유값 - IW



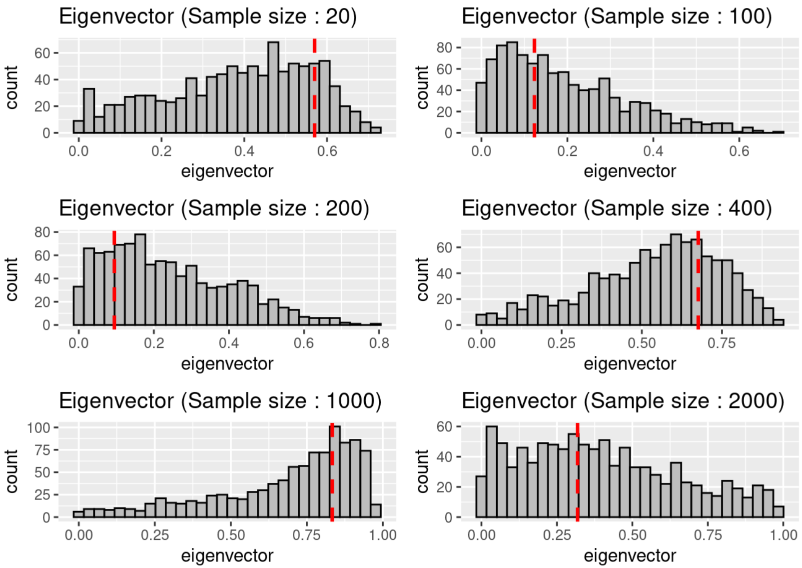
Sample size : 20
Sample size : 100
Sample size : 200
Sample size : 400
Sample size : 1000
Sample size : 2000
lambda_p
lambda_1
2.0
1.5
1.0
0.5
0.0
15
20
25
30
35
40

고유벡터 - SIW



Eigenvector (Sample size : 20)
Eigenvector (Sample size : 100)
Eigenvector (Sample size : 200)
Eigenvector (Sample size : 400)
Eigenvector (Sample size : 1000)
Eigenvector (Sample size : 2000)
count
eigenvector

고유벡터 - IW



Eigenvector (Sample size : 20)
count
eigenvector
Eigenvector (Sample size : 100)
count
eigenvector
Eigenvector (Sample size : 200)
count
eigenvector
Eigenvector (Sample size : 400)
count
eigenvector
Eigenvector (Sample size : 1000)
count
eigenvector
Eigenvector (Sample size : 2000)
count
eigenvector

목차

향후 계획

$$\pi(\Sigma|a,b,H) \propto \frac{1}{|\Sigma|^a[\prod_{i<j}(\lambda_i-\lambda_j)]^b} etr(-\frac{1}{2}\Sigma^{-1}H)$$

- 초모수 a, b, H 가 n, p에 의존하게 설정하여 고유값을 잘 추정하는 사전분포 제안.
-

References I

[1] BERGER, J. O., SUN, D., AND SONG, C. Bayesian analysis of the covariance matrix of a multivariate normal distribution with a new class of priors. *The Annals of Statistics 48*, 4 (2020), 2381–2403.

[2] CHIKUSE, Y. *Statistics on special manifolds*, vol. 1. Springer.

[3] GUPTA, A. K., AND NAGAR, D. K. *Matrix variate distributions*, vol. 104. CRC Press, 1999.

[4] HOFF, P. D. Simulation of the matrix bingham–von mises–fisher distribution, with applications to multivariate and relational data. *Journal of Computational and Graphical Statistics 18*, 2 (2009), 438–456.