Gibbs Sampling with VAR Applications

Frank Schorfheide
University of Pennsylvania

Econ 722 – Part 1

February 5, 2019

Gibbs Sampler

- Suppose the parameter vector θ can be partitioned into $\theta = [\theta_1', \dots, \theta_m']'$.
- For each j it is possible to generate draws of θ_j from the conditional distribution $p(\theta_j|\theta_{-j}, Y)$, where θ_{-j} denotes the vector θ without the partition θ_j.
- For $j = 1, \dots, N$:
  1. Draw θ_1^{i+1} from the density $p(\theta_1|\theta_2^i, \dots, \theta_m^i, Y)$.
  2. Draw θ_2^{i+1} from the density $p(\theta_2|\theta_1^{i+1}, \theta_3^i, \dots, \theta_m^i, Y)$.
  3. $\cdots$
  4. Draw θ_m^{i+1} from the density $p(\theta_m|\theta_1^{i+1}, \dots, \theta_{m-1}^{i+1}, Y)$. $\square$

Application 1: VAR in Deviations from Trend

Consider the following VAR:

$$y_t = \Gamma_0 + \Gamma_1 t + \widetilde{y}_t, \quad \widetilde{y}_t = \Phi_1 \widetilde{y}_{t-1} + \ldots + \Phi_p \widetilde{y}_{t-p} + u_t, \quad u_t \sim iidN(0, \Sigma).$$

Posterior of $(\Phi, \Sigma)|(\Gamma, Y)$

Definitions

- $\Phi = [\Phi_1, \dots, \Phi_p]'$ and $\Gamma = [\Gamma_1', \Gamma_2']'$.
- $\widetilde{Y}(\Gamma)$ is the $T \times n$ matrix with rows $(y_t - \Gamma_0 - \Gamma_1 t)'$.
- $\widetilde{X}(\Gamma)$ is the $T \times (pn)$ matrix with rows $[(y_{t-1} - \Gamma_0 - \Gamma_1(t-1))', \dots, (y_{t-p} - \Gamma_0 - \Gamma_1(t-p))']$.

Posterior of $(\Phi, \Sigma)|(\Gamma, Y)$

- (Conditional) Likelihood Function:

$$p(Y_{1:T}|\Phi,\Sigma,\Gamma,Y_{1-p:0}) \propto |\Sigma|^{-T/2} \times \exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}(\widetilde{Y}(\Gamma)-\widetilde{X}(\Gamma)\Phi)'(\widetilde{Y}(\Gamma)-\widetilde{X}(\Gamma)\Phi)\right]\right\}$$

- Deduce: if prior of $(\Phi,\Sigma)|\Gamma$ is MNIW, then posterior $(\Phi,\Sigma)|(\Gamma,Y)$ is MNIW.

Posterior of $\Gamma | (\Phi, \Sigma, Y)$

- Write

$$\left(I - \sum_{j=1}^{p} \Phi_j L^j \right) (y_t - \Gamma_0 - \Gamma_1 t) = u_t.$$

- Define

$$z_t(\Phi) = \left(I - \sum_{j=1}^{p} \Phi_j L^j \right) y_t, \quad W_t(\Phi) = \left[\left(I - \sum_{j=1}^{p} \Phi_j \right), \left(I - \sum_{j=1}^{p} \Phi_j L^j \right) t \right].$$

- Thus,

$$z_t(\Phi) = W_t(\Phi)\Gamma + u_t.$$

Posterior of $\Gamma|(\Phi, \Sigma, Y)$

- The likelihood function can be re-written as

$$
\begin{aligned}
& p(Y_{1:T}|\Phi, \Sigma, \Gamma, Y_{1-p:0}) \\
& \quad \propto \quad \exp\left\{ -\frac{1}{2}\sum_{t=1}^{T}(z_t(\Phi) - W_t(\Phi)\Gamma)'\Sigma^{-1}(z_t(\Phi) - W_t(\Phi)\Gamma) \right\}.
\end{aligned}
$$

- Deduce: if prior of $\Gamma|(\Phi, \Sigma)$ is MN, then posterior of $\Gamma|(\Phi, \Sigma, Y)$ is MN.

Gibbs Sampler – Algorithm

For $i = 1, \ldots, N$:

1. Draw (Φ^i, Σ^i) from the MNIW distribution of $(\Phi, \Sigma)|(\Gamma^{i-1}, Y)$.
2. Draw Γ^i from the Normal distribution of $\Gamma|(\Phi^i, \Sigma^i, Y)$. $\square$

Application 2: Cointegration

- Consider a VAR(1) model of the form

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + u_t \quad u_t \sim iid\mathcal{N}(0, \Sigma).$$

- The dynamic behavior of the VAR is determined by the eigenvalues of Φ_1 satisfying

$$|\Phi_1 - \lambda I| = 0.$$

- It is interesting to consider specifications for which some eigenvalues are equal to one and the others are less than one in absolute value. These restrictions can be equivalently expressed in terms of the roots of the characteristic polynomial of y_t

$$|\Phi(z)| = |I - \Phi_1 z| = 0$$

We will restrict our attention to models for which the roots are either equal to one, or strictly greater than one in absolute value.

Cointegration

- The VAR can be rewritten in vector error (or equilibrium) correction (VEC) form as

$$\begin{aligned}\Delta y_t &= \Phi_0 + (\Phi_1 - I)y_{t-1} + u_t \\ &= \Pi_0 + \Pi_* y_{t-1} + u_t\end{aligned}$$

- Unit eigenvalues of Φ_1 translate into zero eigenvalues for the matrix Π_*. Hence, Π_* is potentially of reduced rank. We parameterize Π_* as

$$\Pi_* = \alpha\beta',$$

  where α and β are $n \times r$ matrices.

- Note that the re-parameterization is not unique. Let A be a $r \times r$ matrix that is invertible. Then

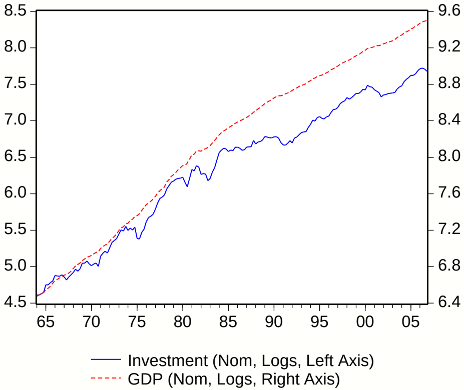
$$\Pi_* = \alpha I_{r\times r}\beta' = \alpha A A^{-1}\beta' = \tilde{\alpha}\tilde{\beta}'.$$

  To obtain a unique re-parameterization, one has to normalize the vector β.

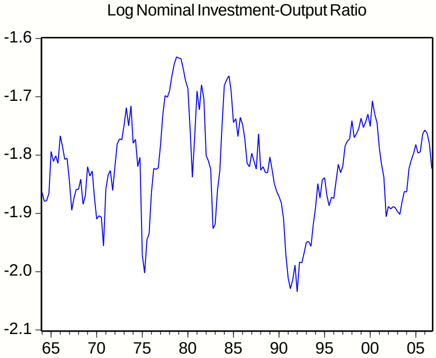
Definition

- Let y_t be integrated of order 1.
- We call y_t cointegrated with cointegrating vectors β_i if $\beta_i' y_t$ can be made stationary by a suitable choice of its initial distribution.
- The cointegration rank is the number of linearly independent cointegrating relations,
- and the space spanned by the cointegration relations is the cointegrating space.

Some Data



8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
9.6
9.2
8.8
8.4
8.0
7.6
7.2
6.8
6.4
65
70
75
80
85
90
95
00
05
Investment (Nom, Logs, Left Axis)
GDP (Nom, Logs, Right Axis)



Log Nominal Investment-Output Ratio
-1.6
-1.7
-1.8
-1.9
-2.0
-2.1
65
70
75
80
85
90
95
00
05

Notation

- In addition to the matrix α of dimension $n \times r$ we consider another matrix $\alpha_\perp$ of full rank and dimension $n \times (n-r)$ such that $\alpha'\alpha_\perp = 0$.
- The matrix $[\alpha, \alpha_\perp]$ has rank n. The matrix $\alpha_\perp$ is not uniquely defined, but whenever it is used the conclusions depend only on the orthogonality property. Define $\beta_\perp$ in similar fashion.
- If $\beta'\alpha$ has full rank then any vector v in $\mathbb{R}^n$ can be decomposed into a vector in the space spanned by $\beta_\perp$, and the space spanned by α.

Granger's Representation Theorem

If $|\Phi(z)| = 0$ implies that $|z| > 1$ or $z = 1$, and the rank of Π_* is $r < n$, then there exist $n \times r$ matrices α and β of rank r such that

$$\Pi_* = \alpha\beta'$$

A necessary and sufficient condition that Δy_t and $\beta' y_t$ can be given initial distributions such that they become $I(0)$ is that

$$\alpha'_\perp \beta_\perp$$

has full rank. In this case the solution y_t has the representation

$$y_t = C \sum_{\tau=1}^{t} (u_\tau + \Pi_0) + \Psi(L)(u_t + \Pi_0) + Cy_0$$

where $C = \beta_\perp(\alpha'_\perp \beta_\perp)^{-1}\alpha'_\perp$. Thus, y_t is a $I(1)$ vector process with cointegrating vectors β.

Additional Lags

- Consider a VAR(p) in error (or equilibrium) correction form:

$$\Delta y_t = \Pi_0 + \Pi_* y_{t-1} + \sum_{j=1}^{p-1} \Pi_i \Delta y_{t-j} + u_t$$

where $\Pi_* = \alpha\beta'$ is of rank $0 \leq r \leq n$.

- If the rank of Π_* is 0, then the model simplifies to a VAR in differences.
- If the rank is equal to n and the roots of the polynomial

$$|\Phi(z)| = \left| I(1-z) - \Pi_* z - \sum_{j=1}^{p-1} \Pi_j z^j (1-z) \right|$$

are outside of the unit circle, then the VAR is stationary.

The Likelihood Function

- Define the $n \times 1$ vector

$$x_t = y_{t-1}$$

- and the $[n(p-1)+1] \times 1$ vector w_t as

$$w_t = [1, \Delta y'_{t-1}, \ldots, \Delta y'_{t-p+1}]'$$

- Moreover, define the matrices

$$\Delta Y = \begin{bmatrix} \Delta y'_1 \\ \vdots \\ \Delta y'_T \end{bmatrix}, \quad X = \begin{bmatrix} x'_1 \\ \vdots \\ x'_T \end{bmatrix}, \quad W = \begin{bmatrix} w'_1 \\ \vdots \\ w'_T \end{bmatrix},$$

$$\Gamma = [\Pi_0, \Pi_1, \ldots, \Pi_{p-1}]'$$

The Likelihood Function

- The conditional density of y_t can be written as

$$\begin{aligned} & p(y_t|Y^{t-1}, \alpha, \beta, \Gamma, \Sigma) \\ & \quad \propto \quad |\Sigma|^{-1/2} \exp\bigg\{ -\frac{1}{2} tr[\Sigma^{-1}(\Delta y_t' - x_t'\beta\alpha' - w_t'\Gamma)' \\ & \qquad \qquad \times(\Delta y_t' - x_t'\beta\alpha' - w_t'\Gamma)] \bigg\} \end{aligned}$$

- Hence, conditional on initial observations

$$\begin{aligned} & p(Y|\alpha, \beta, \Gamma, \Sigma) \\ & \quad = \quad \prod_{t=1}^{T} p(y_t|Y^{t-1}, \alpha, \beta, \Gamma, \Sigma) \\ & \quad \propto \quad |\Sigma|^{-T/2} \exp\bigg\{ -\frac{1}{2} tr[\Sigma^{-1}(\Delta Y - X\beta\alpha' - W\Gamma)' \\ & \qquad \qquad \times(\Delta Y - X\beta\alpha' - W\Gamma)] \bigg\} \end{aligned}$$

Gibbs Sampler

- See Geweke (1996, Journal of Econometrics).
- Geweke uses a shrinkage prior, we simply use the improper prior

$$P(\alpha, \beta, \Gamma, \Sigma) \propto |\Sigma|^{-(n+1)/2}$$

  to illustrate the derivation of the Gibbs sampler.
- Without normalization of the cointegration vector, the model is not identified.
- Recall that β' is an $r \times n$ matrix composed of the cointegration vectors. We will use the normalization

$$\beta' = [\underbrace{I}_{r \times r}, \underbrace{\tilde{\beta}'}_{r \times (n-r)}]$$

Gibbs Sampler

Iterate over the following conditionals:

- $p(\Sigma|\alpha', \beta, \Gamma, Y)$
- $p(\Gamma|\alpha', \beta, \Sigma, Y)$
- $p(\alpha'|\beta, \Gamma, \Sigma, Y)$
- $p(\beta|\alpha', \Gamma, \Sigma, Y)$

Gibbs Sampler: $p(\Sigma|\alpha', \beta, \Gamma, Y)$

- Define

$$S(\alpha', \beta, \Gamma) = (\Delta Y - X\beta\alpha' - W\Gamma)'(\Delta Y - X\beta\alpha' - W\Gamma)$$

  and $\nu = T$.
- Then

$$\Sigma|\alpha', \beta, \Gamma, Y \sim \mathcal{IW}\Big(S(\alpha', \beta, \Gamma), T\Big)$$

Gibbs Sampler: $p(\Gamma|\alpha', \beta, \Sigma, Y)$

- Define

$$\widetilde{Y}(\alpha', \beta) = \Delta Y - X\beta\alpha'$$

- Then

$$\widetilde{Y}(\alpha', \beta) = W\Gamma + U$$

- Define

$$\widehat{\Gamma} = (W'W)^{-1}W'\widetilde{Y}(\alpha', \beta)$$

- and deduce

$$\Gamma|\alpha', \beta, \Sigma, Y \sim \mathcal{N}\left(\widehat{\Gamma}, \Sigma \otimes (W'W)^{-1}\right)$$

Gibbs Sampler: $p(\alpha'|\beta, \Gamma, \Sigma, Y)$

- Define

$$\widetilde{Z}(\Gamma) = \Delta Y - W\Gamma \quad \text{and} \quad \widetilde{X}(\beta) = X\beta$$

- Then

$$\widetilde{Z}(\Gamma) = \widetilde{X}(\beta)\alpha' + U$$

- Define

$$\widehat{\alpha'} = \left(\widetilde{X}(\beta)'\widetilde{X}(\beta)\right)^{-1}\widetilde{X}(\beta)'\widetilde{Z}(\Gamma)$$

- and deduce

$$\alpha'|\beta, \Gamma, \Sigma, Y \sim \mathcal{N}\left(\widehat{\alpha'}, \Sigma \otimes \left(\widetilde{X}(\beta)'\widetilde{X}(\beta)\right)^{-1}\right)$$

Gibbs Sampler: $p(\beta|\alpha', \Gamma, \Sigma, Y)$

- Recall that we have to normalize $\beta' = [I, \tilde{\beta}']$. We will also partition $X = [X_1, X_2]$ such that the partitions of X conform with the partitions β.
- Then we have

$$\begin{aligned} \Delta Y &= X\beta\alpha' + W\Gamma + U \\ &= X_1\alpha' + X_2\tilde{\beta}\alpha' + W\Gamma + U \end{aligned}$$

- Define $Z(\alpha', \Gamma) = \Delta Y - X_1\alpha' - W\Gamma$ and write

$$Z = X_2\tilde{\beta}\alpha' + U.$$

- Post-multiply the above equation by the matrix

$$C = [\alpha(\alpha'\alpha)^{-1}, \alpha_\perp].$$

Gibbs Sampler: $p(\beta|\alpha', \Gamma, \Sigma, Y)$

- We obtain the seemingly unrelated regression

$$
\begin{aligned}
ZC &= X_2\tilde{\beta}\alpha' C + UC \\
\left[Z\alpha(\alpha'\alpha)^{-1},\ Z\alpha_{\perp} \right] &= \left[X_2\tilde{\beta},\ 0 \right] + \left[U\alpha(\alpha'\alpha)^{-1}, U\alpha_{\perp} \right] \\
[\tilde{Z}_1,\ \tilde{Z}_2] &= X_2[\tilde{\beta},\ 0] + [\tilde{U}_1,\ \tilde{U}_2]
\end{aligned}
$$

- Let $\tilde{\Sigma} = C'\Sigma C$ and partition

$$
\tilde{\Sigma} = \left[\begin{array}{cc} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{12} \\ \tilde{\Sigma}_{21} & \tilde{\Sigma}_{22} \end{array} \right]
$$

Gibbs Sampler: $p(\beta|\alpha', \Gamma, \Sigma, Y)$

- We will use the following fact about multivariate normal distributions. Suppose $\tilde{u} \sim \mathcal{N}(0, \tilde{\Sigma})$ then,

$$\begin{aligned} \tilde{u}_2 &\sim N\left(0, \tilde{\Sigma}_{22}\right) \\ \tilde{u}_1|\tilde{u}_2 &\sim N\left(\Sigma_{12}\Sigma_{22}^{-1}u_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right) \end{aligned}$$

- Let $\tilde{\Sigma}_{11.22} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. This implies that

$$\begin{aligned} & p(\beta|\alpha', \Gamma, \Sigma, Y) \\ & \propto \quad |\tilde{\Sigma}_{22}|^{-T/2} \exp\left\{ -\frac{1}{2} tr\left[\tilde{\Sigma}_{22}^{-1}\tilde{Z}_2'\tilde{Z}_2\right]\right\} \\ & \quad \times |\tilde{\Sigma}_{11.22}|^{-T/2} \exp\left\{ -\frac{1}{2} tr\left[\tilde{\Sigma}_{11.22}^{-1}(\tilde{Z}_1 - X_2\tilde{\beta} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{Z}_2)' \right.\right. \\ & \quad \left.\left. \times (\tilde{Z}_1 - X_2\tilde{\beta} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{Z}_2)\right]\right\} \end{aligned}$$

Gibbs Sampler: $p(\beta|\alpha', \Gamma, \Sigma, Y)$

- Hence, define

$$\bar{\tilde{\beta}} = (X_2'X_2)^{-1} X_2'(\tilde{Z}_1 - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{Z}_2)$$
$$V_{\tilde{\beta}} = \tilde{\Sigma}_{11.22} \otimes (X_2'X_2)^{-1}$$

- and we obtain

$$\beta|\alpha', \Gamma, \Sigma, Y \sim N\left(\bar{\tilde{\beta}}, V_{\tilde{\beta}}\right)$$

Gibbs Sampler – Some Intuition

- Suppose we iterate over

$$p(\theta|\phi), \quad p(\phi|\theta).$$

- Define marginals

$$p(\theta) = \int_{\Phi} p(\theta|\phi)p(\phi)d\phi, \quad p(\phi) = \int_{\Theta} p(\phi|\tilde{\theta})p(\tilde{\theta})d\tilde{\theta}.$$

- Combine:

$$p(\theta) \;=\; \int_{\Phi} p(\theta|\phi)\left[\int_{\Theta} p(\phi|\tilde{\theta})p(\tilde{\theta})d\tilde{\theta}\right] d\phi = \int_{\Theta}\left[\int_{\Phi} p(\theta|\phi)p(\phi|\tilde{\theta})d\phi\right] p(\tilde{\theta})d\tilde{\theta}$$

- Define Markov transition kernel:

$$K(\theta|\tilde{\theta}) = \int_{\Phi} p(\theta|\phi)p(\phi|\tilde{\theta})d\phi$$

Gibbs Sampler – Some Intuition

- Recall Markov transition kernel:

$$K(\theta|\tilde{\theta}) = \int_{\Phi} p(\theta|\phi)p(\phi|\tilde{\theta})d\phi$$

- Note that $p(\theta)$ is a fixed point of the mapping $M[\cdot]$:

$$p(\theta) = \int K(\theta|\tilde{\theta})p(\tilde{\theta})d\tilde{\theta} = M[p(\tilde{\theta})]$$

- Questions (see Tanner and Wong (1987) for answers):
  - Is the fixed point unique? Yes
  - Is $M[\cdot]$ a contraction mapping? Yes

Some Regularity Conditions

- $K(\theta|\tilde{\theta})$ is uniformly bounded and equicontinuous in θ.
- For any $\theta_0 \in \Theta$ there is a neighborhood $U(\theta_0)$ such that $K(\theta|\tilde{\theta}) > 0$ for all $\theta, \tilde{\theta} \in U(\theta_0)$.

Some Notation

- For a function $f(\theta)$ let $\|f\| = \int |f(\theta)| d\theta$.
- Recall the map $M[f] = \int K(\theta|\tilde{\theta}) f(\tilde{\theta}) d\tilde{\theta}$. Note that $M[\cdot]$ can be applied to a large class of functions $f(\cdot)$ (not just densities).

Useful Results

Lemma 1

Every fixed point of $M[\cdot]$ must be continuous.

- Let $p_*(\theta)$ be a fixed point of $M[\cdot]$.
- Consider

$$\begin{aligned}
&\lim_{\theta_1 \longrightarrow \theta_0} |p_*(\theta_1) - p_*(\theta_0)| \\
&= \lim_{\theta_1 \longrightarrow \theta_0} \left| \int K(\theta_1|\tilde{\theta}) p_*(\tilde{\theta}) d\tilde{\theta} - \int K(\theta_0|\tilde{\theta}) p_*(\tilde{\theta}) d\tilde{\theta} \right| \\
&\leq \lim_{\theta_1 \longrightarrow \theta_0} \int \left| K(\theta_1|\tilde{\theta}) - K(\theta_0|\tilde{\theta}) \right| p_*(\tilde{\theta}) d\tilde{\theta} \\
&= \int \left[\lim_{\theta_1 \longrightarrow \theta_0} \left| K(\theta_1|\tilde{\theta}) - K(\theta_0|\tilde{\theta}) \right| \right] p_*(\tilde{\theta}) d\tilde{\theta} \\
&= 0
\end{aligned}$$

- The second-to-last equality follows from the assumptions.

Useful Results

Lemma 2

$\|M[|f|]\| = \|f\|$

- Note that

$$\begin{aligned}
\|M[|f|]\| &= \int_{\Theta}\left[\int_{\tilde{\Theta}} K(\theta|\tilde{\theta})|f(\tilde{\theta})|d\tilde{\theta}\right] d\theta \\
&= \int_{\tilde{\Theta}}\left[\int_{\Theta} K(\theta|\tilde{\theta})d\theta\right] |f(\tilde{\theta})|d\tilde{\theta} \\
&= \int_{\tilde{\Theta}} |f(\tilde{\theta})|d\tilde{\theta} \\
&= \|f\|
\end{aligned}$$

Useful Results

Lemma 3

$\|M[f]\| \leq \|f\|$

- Note that

$$\begin{aligned} \|M[f]\| &= \int |M[f]| d\theta \\ &\leq \int M[|f|] d\theta \\ &= \|M[|f|]\| \\ &= \|f\| \end{aligned}$$

Useful Results

Lemma 4

Let $f^+ = f\{f \geq 0\}$ and $f^- = (-f)\{f < 0\}$. If f is such that neither f^+ nor f^- are identical to zero, then $\|M[f]\| < \|f\|$.

- Recall that $M[f] = \int K(\theta|\tilde{\theta})f(\tilde{\theta})d\tilde{\theta}$.
- Now consider

$$M[|f|] = M[f^+ + f^-] = M[f^+] + M[f^-], \quad |M[f]| = |M[f^+] - M[f^-]|.$$

- Note: $supp(f^+) \subset supp(M[f^+])$ and $supp(f^-) \subset supp(M[f^-])$.
- Deduce $supp(M[f^+])$ and $supp(M[f^-])$ overlap.
- In turn, $|M[f^+] - M[f^-]| < M[f^+ + f^-] = M[f^+] + M[f^-]$.
- Thus, $\|M[f]\| < \|M[|f|]\| = \|f\|$ (Lemma 2).

Uniqueness of Fixed Point

Uniqueness

p_* is the only density that satisfies $p_* = M[p_*]$.

- Suppose (to the contrary) $p_{**} = M[p_{**}]$ and define $f = p_* - p_{**}$.
- Then $M[f] = M[p_* - p_{**}] = p_* - p_{**} = f$ and f is a fixed point.
- f must be continuous (Lemma 1).
- Since $\int f(\theta)d\theta = 0$ and $f(\theta) \neq 0$ neither f^+ nor f^- can be zero.
- Thus, $\|M[f]\| < \|f\|$ (Lemma 4), which contradicts that f is a fixed point.

Contraction Mapping

We want that $\|p_{(s+1)} - p_*\| < \|p_{(s)} - p_*\|$, where $p_{(s+1)} = M[p_{(s)}]$.

- It is straightforward to show the weaker result: $\|p_{(s+1)} - p_*\| \leq \|p_{(s)} - p_*\|$.
- Let $f = p_{(s)} - p_*$ such that $M[f] = p_{(s+1)} - p_*$.
- Desired result follows from Lemma 3 which states that $\|M[f]\| \leq \|f\|$.
- One can use arguments similar to those on the previous slide to turn the weak inequality into a strict inequality.

Convergence Rates

- Suppose that the starting value $p_{(0)}(\theta)$ satisfies $\sup_\theta \; p_{(0)}(\theta)/p_*(\theta) < \infty$.
- Then there exists a constant $\alpha \in (0, 1)$ such that

$$\|p_{(s)} - p_*\| \leq \alpha^s \|p_{(0)} - p_*\|$$

- See Tanner and Wong (1987).

Summary

- Let $p(\theta)$ be a normalized probability density. Define the mapping

$$M[p(\theta)] = \int K(\theta|\tilde{\theta}, Y)p(\tilde{\theta})d\tilde{\theta}$$

  $M[\cdot]$ maps a density $p(\theta)$ into a density $p'(\theta)$.

- We are interested in applying the mapping iteratively: Let $p^i(\theta) = M[p^{i-1}(\theta)]$.
- The mapping is constructed such that the fixed point corresponds to the posterior of interest.
- Under suitable regularity conditions
  1. The fixed point $p_*(\theta)$ of the mapping $M[\cdot]$ is unique.
  2. The mapping $M[\cdot]$ is a contraction mapping and the sequence of densities $\{p^i(\theta)\}_{i=0}^{\infty}$ converges to the fixed point $p_*(\theta)$

$$\int |p^i(\theta) - p_*(\theta)|d\theta \longrightarrow 0$$

  as $i \longrightarrow \infty$. $\square$

Gibbs Sampler

- For $i = 1, \ldots, N$:
  1. Draw ϕ^{i+1} from the density $p(\phi|\theta^i)$.
  2. Draw θ^{i+1} from the density $p(\theta|\phi^{i+1})$.
- It turns out that for $s > \bar{S}$ the marginal distribution of the draws (θ^i, ϕ^i) is approximately equal to the target distribution $p(\theta, \phi)$.
- However, the sequence of draws is serially correlated!
- Gibbs sampler creates a Markov chain. It belongs to the class of Markov chain Monte Carlo (MCMC) procedures.

Gibbs Sampler

- Suppose the parameter vector θ can be partitioned into $\theta = [\theta_1', \ldots, \theta_m']'$.
- For each j it is possible to generate draws of θ_j from the conditional distribution $p(\theta_j|\theta_{-j}, Y)$, where θ_{-j} denotes the vector θ without the partition θ_j.
- For $j = 1, \ldots, N$:
  1. Draw θ_1^{i+1} from the density $p(\theta_1|\theta_2^i, \ldots, \theta_m^i, Y)$.
  2. Draw θ_2^{i+1} from the density $p(\theta_2|\theta_1^{i+1}, \theta_3^i, \ldots, \theta_m^i, Y)$.
  3. $\cdots$
  4. Draw θ_m^{i+1} from the density $p(\theta_m|\theta_1^{i+1}, \ldots, \theta_{m-1}^{i+1}, Y)$. $\square$

Ergodicity

- A stationary process $\{\theta^i\}$ is said to be ergodic, if for any two bounded and measurable functions $f(\cdot)$ and $g(\cdot)$:

$$\begin{aligned} \lim_{n \longrightarrow \infty} & \left| \mathbb{E}\big[f(\theta^i, \ldots, \theta^{i+k}) g(\theta^{i+n}, \ldots, \theta^{i+n+l}) \big] \right| \\ - & \left| \mathbb{E}\big[f(\theta^i, \ldots, \theta^{i+k}) \big] \right| \cdot \left| \mathbb{E}\big[g(\theta^{i+n}, \ldots, \theta^{i+n+l}) \big] \right| = 0. \end{aligned}$$

- If $\{\theta^i\}$ is strictly stationary and ergodic with $\mathbb{E}[|h(\theta)|] < \infty$, then

$$\frac{1}{N} \sum_{i=1}^{N} h(\theta^i) \xrightarrow{a.s.} \mathbb{E}[h(\theta)].$$

A Sufficient Condition for Ergodicity

Suppose that for every $\theta \in \Theta$ and every $A \subseteq \Theta$

$$\int_A p(\theta|Y)d\theta > 0 \quad \text{implies} \quad \int_A K(\tilde{\theta}|\theta)d\tilde{\theta} > 0$$

then the transition kernel of the Gibbs sampler is ergodic. (Geweke, 2005, Corollary 4.5.1)

Another Sufficient Condition for Ergodicity

Suppose that the following three conditions are satisfied:

- For all θ with $p(\theta|Y) > 0$ there exists an open neighborhood $N_\delta(\theta)$ such that for all $\tilde{\theta} \in N_\delta(\theta)$ $p(\tilde{\theta}|Y) > 0$.
- For every point $\tilde{\theta} \in \Theta$ and each block b of the Gibbs sampler, there exists an open neighborhood $N_\delta(\tilde{\theta}_{-b})$ of $\tilde{\theta}_{-b}$ and a bounded function $c(\tilde{\theta}_{-b})$ such that for all $\theta_{-b} \in N_\delta(\tilde{\theta}_{-b})$

$$\int_{\Theta(b)} p(\tilde{\theta}_{<b}, \theta_b, \tilde{\theta}_{>b}) d\theta_b \leq c(\tilde{\theta}_{-b})$$

- Θ is connected.

Then the transition kernel of the Gibbs sampler is ergodic. (Geweke, 2005, Theorem 4.5.4)

Gibbs Sampler

- For large N we obtain dependent draws from the posterior distribution of θ. It is common practice to discard the initial draws.
- Approximate the mean and covariance matrix of θ by Monte Carlo averages:

$$\widehat{\mathbb{E}[\theta]} = \frac{1}{N - N_0} \sum_{i=N_0+1}^{N} h(\theta^i) \stackrel{a.s.}{\longrightarrow} \mathbb{E}[h(\theta)|Y]$$

provided $\mathbb{E}\big[|\theta(\theta)|\big|Y\big] < \infty$.

- Stronger regularity conditions are required to obtain a Central Limit Theorem (CLT)

$$\sqrt{N - N_0}\big(\widehat{\mathbb{E}}[\theta|Y] - \mathbb{E}[\theta|Y]\big) \Longrightarrow N(0, V)$$

- A CLT facilitates the computation of numerical standard errors for Monte Carlo approximations.