To Center or Not to Center, That is Not the Question: An Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency

Xiao-Li Meng

Department of Statistics, Harvard University

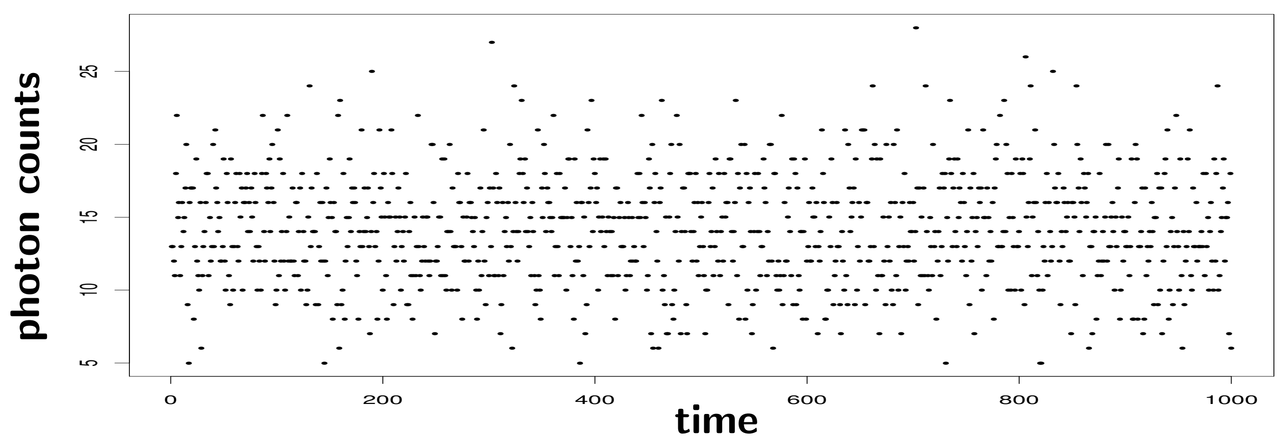
Joint work with Yaming Yu of UC Irvine.

We thank David van Dyk and Vinay Kashyap for stimulating discussions.

Astrophysics: Source Intensity Variations

- Poisson variation of the counts given the intensity.
- Variation of the intensity itself.
  - X-ray flare
  - Binary systems
  - Gradual cooling

The isolated neutron star/quark star candidate RX J1856.5-3754 observed by Chandra HRC (exposure time 55476 seconds, divided into 1000 bins).



A Parameter-Driven Poisson Time Series Model

$$\theta = (\beta_0 \quad \beta_1 \quad \rho \quad \delta)$$

baseline (β_0) **trend** (β_1) **autocorr.** (ρ) **residual s.d.** (δ)

Y_{obs} : **counts observed**

Y_{mis} : **depends on the augmentation scheme**

$$Y_t|(\xi_t, \beta) \overset{ind}{\sim} Pois\left(d_t e^{\beta_0+\beta_1 t+\xi_t}\right);$$

$$\xi_t|(\xi_{<t}, \beta, \rho, \delta) \sim N(\rho\xi_{t-1},\ \delta^2).$$

- Y_t: **counts in bin** t, $\quad t = 1, \ldots, T$;
- d_t: **width (e.g., in seconds) of bin** t;
- $\xi = \{\xi_t\}$ **is a stationary AR(1) process;**
  $\xi_t \sim N(0,\ \tau^2)$, **where** $\tau^2 = \delta^2/(1-\rho^2)$.

Posterior Simulation: The Standard Gibbs Sampler

$$Y_t \sim Pois\left(d_t e^{\beta_0+\beta_1 X_t+\xi_t}\right), \quad \xi_t|\xi_{<t} \sim N(\rho\xi_{t-1},\ \delta^2); \quad p(\beta,\rho,\tau) \propto 1$$

1. $\xi|(\beta,\ \rho,\ \delta)$: draw ξ, the missing data.

   Difficult to update all ξ's simultaneously, so update $\xi_t|(\xi_{t-1},\ \xi_{t+1})$ in turn.

2. $\beta|(\xi,\ \rho,\ \delta)$, or $\beta|\xi$

   Equivalent to posterior sampling of a Poisson GLM. Need an M–H move.

3. $(\rho,\ \delta)|(\xi,\ \beta)$, or $(\rho,\ \delta)|\xi$

   Equivalent to Bayesian fitting of an AR(1) model:

$$\xi_t = \rho\xi_{t-1} + N(0,\ \delta^2), \quad t = 1, \ldots, T.$$

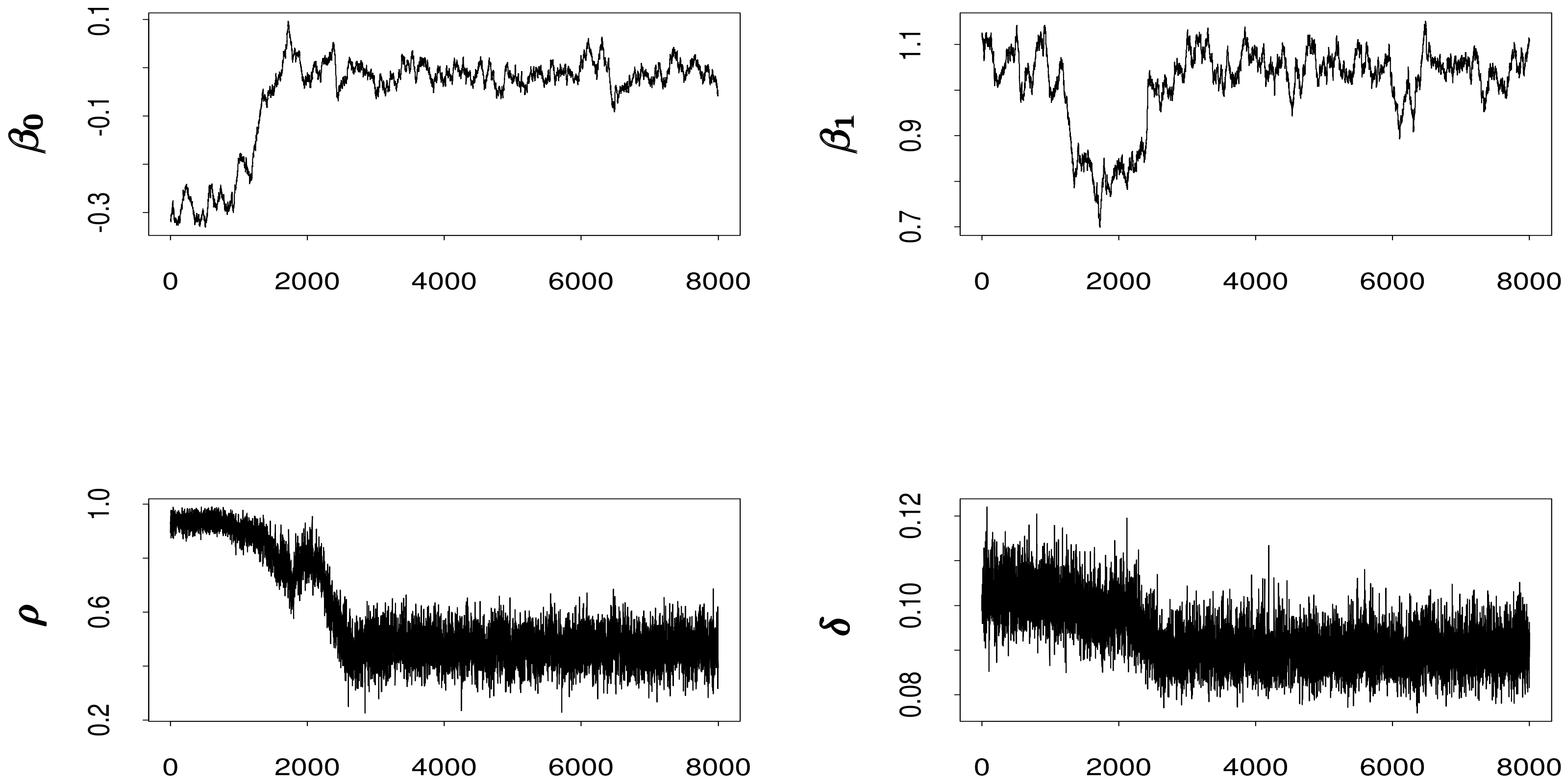
Poor Performance of the Standard Gibbs Sampler

A simulation:

- **Counts are generated according to the correct model**
  - $T = 200$, $d_t = 5000$, **and** $X_t = t/T$.
  - **Parameter values:** $(\beta_0, \beta_1, \rho, \delta) = (0, 1, 0.5, 0.1)$.
- **Counts are in the order of thousands.**

16000 MCMC draws (excluding a burn-in period of 4000), starting from the true parameter values and keeping every other draw:

Poor Performance of the Standard Gibbs Sampler



Data Augmentation – Definition

- **Model of interest:**

$$p(\theta|Y_{obs}) \propto p(Y_{obs}|\theta)p(\theta)$$

Data Augmentation – Definition

- **Model of interest:**

$$p(\theta|Y_{obs}) \propto p(Y_{obs}|\theta)p(\theta)$$

- **Data Augmented (DA) Model:**

$$p(Y_{mis}, Y_{obs}|\theta) = p(Y_{mis}|Y_{obs};\theta)p(Y_{obs}|\theta)$$

- **So we can perform Gibbs Sampler (or other MCMC algorithms)**

$$Y_{mis}|(\theta, Y_{obs}) \longleftrightarrow \theta|(Y_{mis}, Y_{obs})$$

Data Augmentation – Definition

- **Model of interest:**

$$p(\theta|Y_{obs}) \propto p(Y_{obs}|\theta)p(\theta)$$

- **Data Augmented (DA) Model:**

$$p(Y_{mis}, Y_{obs}|\theta) = p(Y_{mis}|Y_{obs};\theta)p(Y_{obs}|\theta)$$

- **We can even introduce working parameter α:**

$$p(Y_{mis}, Y_{obs}|\theta, \alpha) = p(Y_{mis}|Y_{obs};\theta, \alpha)p(Y_{obs}|\theta)$$

Data Augmentation – Definition

- **Model of interest:**

$$p(\theta|Y_{obs}) \propto p(Y_{obs}|\theta)p(\theta)$$

- **Data Augmented (DA) Model:**

$$p(Y_{mis}, Y_{obs}|\theta) = p(Y_{mis}|Y_{obs};\theta)p(Y_{obs}|\theta)$$

- **We can even introduce working parameter α:**

$$p(Y_{mis}, Y_{obs}|\theta, \alpha) = p(Y_{mis}|Y_{obs};\theta, \alpha)p(Y_{obs}|\theta)$$

- **We can seek Efficient DA via Conditional DA (Meng and van Dyk 1997, 1999; van Dyk and Meng, 2001): finding an "optimal" α.**

Data Augmentation – Definition

- Model of interest:

$$p(\theta|Y_{obs}) \propto p(Y_{obs}|\theta)p(\theta)$$

- Data Augmented (DA) Model:

$$p(Y_{mis}, Y_{obs}|\theta) = p(Y_{mis}|Y_{obs};\theta)p(Y_{obs}|\theta)$$

- We can even introduce working parameter α:

$$p(Y_{mis}, Y_{obs}|\theta, \alpha) = p(Y_{mis}|Y_{obs};\theta, \alpha)p(Y_{obs}|\theta)$$

- We can seek Efficient DA via Conditional DA (Meng and van Dyk 1997, 1999; van Dyk and Meng, 2001): finding an "optimal" α.
- We can seek "optimal working prior" on α as in Marginal DA (or PX-DA of Liu and Wu 1999).

Data Augmentation – Definition

- **Model of interest:**

$$p(\theta|Y_{obs}) \propto p(Y_{obs}|\theta)p(\theta)$$

- **Data Augmented (DA) Model:**

$$p(Y_{mis}, Y_{obs}|\theta) = p(Y_{mis}|Y_{obs};\theta)p(Y_{obs}|\theta)$$

- **We can even introduce working parameter α:**

$$p(Y_{mis}, Y_{obs}|\theta, \alpha) = p(Y_{mis}|Y_{obs};\theta, \alpha)p(Y_{obs}|\theta)$$

- **We can seek Efficient DA via Conditional DA (Meng and van Dyk 1997, 1999; van Dyk and Meng, 2001): finding an "optimal" α.**

- **We can seek "optimal working prior" on α as in Marginal DA (Liu and Wu 1999).**

- **Can we use more than one DA? Which ones? How?**

A Toy Example

- **Model:** $Y = \theta + Z + \epsilon, \quad Z \sim N(0, \tau^2), \quad \epsilon \sim N(0,1), \quad Z \perp \epsilon$
- **Assuming** $p(\theta) \propto 1 \quad \Longrightarrow \quad \theta | Y \sim N(Y, 1 + \tau^2) \longleftarrow$ **our target**

A Toy Example

- **Model:** $Y = \theta + Z + \epsilon, \quad Z \sim N(0, \tau^2), \quad \epsilon \sim N(0,1), \quad Z \perp \epsilon$
- **Assuming** $p(\theta) \propto 1 \quad \Longrightarrow \quad \theta | Y \sim N(Y, 1+\tau^2)$ $\longleftarrow$ **our target**
- **SA:** $(Y_{obs}, Y_{mis}) = (Y, \theta + Z)$ – **Sufficiency:** $Y_{obs}|(\theta, Y_{mis}) \sim N(Y_{mis}, 1)$

$$Y_{mis}|(\theta, Y_{obs}) \sim N\left(\frac{\theta + \tau^2 Y}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right),$$

$$\theta|(Y_{mis}, Y_{obs}) \sim N\left(Y_{mis},\ \tau^2\right).$$

A Toy Example

- **Model:** $Y = \theta + Z + \epsilon, \quad Z \sim N(0, \tau^2), \quad \epsilon \sim N(0,1), \quad Z \perp \epsilon$
- **Assuming** $p(\theta) \propto 1 \quad \Longrightarrow \quad \theta | Y \sim N(Y, 1+\tau^2) \longleftarrow$ **our target**
- **SA:** $(Y_{obs}, Y_{mis}) = (Y, \theta + Z)$ – **Sufficiency:** $Y_{obs} | (\theta, Y_{mis}) \sim N(Y_{mis}, 1)$

$$Y_{mis} | (\theta, Y_{obs}) \sim N\left(\frac{\theta + \tau^2 Y}{1+\tau^2}, \ \frac{\tau^2}{1+\tau^2}\right),$$
$$\theta | (Y_{mis}, Y_{obs}) \sim N\left(Y_{mis}, \ \tau^2\right) .$$

- **AA:**$(Y_{obs}, \tilde{Y}_{mis}) = (Y, Z)$ – **Ancillarity:** $\tilde{Y}_{mis} | \theta \sim N(0, \ \tau^2)$.

$$\tilde{Y}_{mis} | (\theta, Y_{obs}) \sim N\left(\frac{\tau^2 (Y - \theta)}{1+\tau^2}, \ \frac{\tau^2}{1+\tau^2}\right),$$
$$\theta | (\tilde{Y}_{mis}, Y_{obs}) \sim N(Y - \tilde{Y}_{mis}, \ 1).$$

A Toy Example

- **Model:** $Y = \theta + Z + \epsilon, \quad Z \sim N(0, \tau^2), \quad \epsilon \sim N(0,1), \quad Z \perp \epsilon$
- **Assuming** $p(\theta) \propto 1 \quad \Longrightarrow \quad \theta | Y \sim N(Y, 1 + \tau^2) \longleftarrow$ **our target**
- **SA:** $(Y_{obs}, Y_{mis}) = (Y, \theta + Z)$ – **Sufficiency:** $Y_{obs} | (\theta, Y_{mis}) \sim N(Y_{mis}, 1)$

$$Y_{mis} | (\theta, Y_{obs}) \sim N\left(\frac{\theta + \tau^2 Y}{1 + \tau^2}, \; \frac{\tau^2}{1 + \tau^2}\right),$$

$$\theta | (Y_{mis}, Y_{obs}) \sim N\left(Y_{mis}, \; \tau^2\right).$$

- **AA:Treat** $(Y_{obs}, \tilde{Y}_{mis}) = (Y, Z)$ – **Ancillarity:** $\tilde{Y}_{mis} | \theta \sim N(0, \; \tau^2)$.

$$\tilde{Y}_{mis} | (\theta, Y_{obs}) \sim N\left(\frac{\tau^2 (Y - \theta)}{1 + \tau^2}, \; \frac{\tau^2}{1 + \tau^2}\right),$$

$$\theta | (\tilde{Y}_{mis}, Y_{obs}) \sim N(Y - \tilde{Y}_{mis}, \; 1).$$

- **One-to-one mapping:** $\tilde{Y}_{mis} = M_\theta(Y_{mis}) = Y_{mis} - \theta$.

A Toy Example – Stochastic Recursions

- **SA:** $Y_{mis}|\theta \sim N\left(\frac{\theta+\tau^2 Y}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad \theta|Y_{mis} \sim N\left(Y_{mis},\ \tau^2\right)$.

$$\theta^{(t+1)} = \frac{\tau^2}{1+\tau^2}Y + \frac{1}{1+\tau^2}\theta^{(t)} + \sqrt{\frac{\tau^4+2\tau^2}{1+\tau^2}}\delta_1^{(t)}, \qquad \delta_1^{(t)}\ i.i.d\ \sim N(0,1)$$

A Toy Example – Stochastic Recursions

- **SA:** $Y_{mis}|\theta \sim N\left(\frac{\theta+\tau^2 Y}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad \theta|Y_{mis} \sim N\left(Y_{mis},\ \tau^2\right)$.

$$\theta^{(t+1)} = \frac{\tau^2}{1+\tau^2}Y + \frac{1}{1+\tau^2}\theta^{(t)} + \sqrt{\frac{\tau^4+2\tau^2}{1+\tau^2}}\delta_1^{(t)}, \qquad \delta_1^{(t)}\ i.i.d\ \sim N(0,1)$$

- **Convergence rate** $r_{SA} = \frac{1}{1+\tau^2}$**: fast when** τ^2 **is large.**

A Toy Example – Stochastic Recursions

- **SA:** $Y_{mis}|\theta \sim N\left(\frac{\theta+\tau^2 Y}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad \theta|Y_{mis} \sim N\left(Y_{mis},\ \tau^2\right).$

$$\theta^{(t+1)} = \frac{\tau^2}{1+\tau^2}Y + \frac{1}{1+\tau^2}\theta^{(t)} + \sqrt{\frac{\tau^4+2\tau^2}{1+\tau^2}}\delta_1^{(t)}, \qquad \delta_1^{(t)}\ i.i.d\ \sim N(0,1)$$

- **Convergence rate** $r_{SA} = \frac{1}{1+\tau^2}$**: fast when** τ^2 **is large.**
- **AA:** $\tilde{Y}_{mis}|\theta \sim N\left(\frac{\tau^2(Y-\theta)}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad \theta|\tilde{Y}_{mis} \sim N(Y-\tilde{Y}_{mis},\ 1).$

$$\theta^{(t+1)} = \frac{1}{1+\tau^2}Y + \frac{\tau^2}{1+\tau^2}\theta^{(t)} + \sqrt{\frac{2\tau^2+1}{1+\tau^2}}\delta_2^{(t)}, \qquad \delta_2^{(t)}\ i.i.d\ \sim N(0,1)$$

A Toy Example – Stochastic Recursions

- **SA:** $Y_{mis}|\theta \sim N\left(\frac{\theta+\tau^2 Y}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad \theta|Y_{mis} \sim N\left(Y_{mis},\ \tau^2\right).$

$$\theta^{(t+1)} = \frac{\tau^2}{1+\tau^2}Y + \frac{1}{1+\tau^2}\theta^{(t)} + \sqrt{\frac{\tau^4+2\tau^2}{1+\tau^2}}\delta_1^{(t)}, \qquad \delta_1^{(t)}\ i.i.d\ \sim N(0,1)$$

- **Convergence rate** $r_{SA} = \frac{1}{1+\tau^2}$**: fast when** τ^2 **is large.**
- **AA:** $\tilde{Y}_{mis}|\theta \sim N\left(\frac{\tau^2(Y-\theta)}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad \theta|\tilde{Y}_{mis} \sim N(Y-\tilde{Y}_{mis},\ 1).$

$$\theta^{(t+1)} = \frac{1}{1+\tau^2}Y + \frac{\tau^2}{1+\tau^2}\theta^{(t)} + \sqrt{\frac{2\tau^2+1}{1+\tau^2}}\delta_2^{(t)}, \qquad \delta_2^{(t)}\ i.i.d\ \sim N(0,1)$$

- **Convergence rate** $r_{AA} = \frac{\tau^2}{1+\tau^2}$**: fast when** τ^2 **is small.**

A Toy Example – Stochastic Recursions

- **SA:** $Y_{mis}|\theta \sim N\left(\frac{\theta+\tau^2 Y}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right),\quad \theta|Y_{mis} \sim N\left(Y_{mis},\ \tau^2\right).$

$$\theta^{(t+1)} = \frac{\tau^2}{1+\tau^2}Y + \frac{1}{1+\tau^2}\theta^{(t)} + \sqrt{\frac{\tau^4+2\tau^2}{1+\tau^2}}\delta_1^{(t)}, \qquad \delta_1^{(t)}\ i.i.d\ \sim N(0,1)$$

- **Convergence rate** $r_{SA} = \frac{1}{1+\tau^2}$: **fast when** τ^2 **is large.**
- **AA:** $\tilde{Y}_{mis}|\theta \sim N\left(\frac{\tau^2(Y-\theta)}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right),\quad \theta|\tilde{Y}_{mis} \sim N(Y-\tilde{Y}_{mis},\ 1).$

$$\theta^{(t+1)} = \frac{1}{1+\tau^2}Y + \frac{\tau^2}{1+\tau^2}\theta^{(t)} + \sqrt{\frac{2\tau^2+1}{1+\tau^2}}\delta_2^{(t)}, \qquad \delta_2^{(t)}\ i.i.d\ \sim N(0,1)$$

- **Convergence rate** $r_{AA} = \frac{\tau^2}{1+\tau^2}$: **fast when** τ^2 **is small.**
- **Direct combination leads to (counting** $\theta^{(t)} \to \theta^{(t+2)}$ **as one iteration)**

$$r_{SA\times AA} = \frac{\tau^2}{(1+\tau^2)^2}$$

But can we do better?

An Interweaving Scheme: ASIS

- **SA:** $(A)\ \boldsymbol{Y_{mis}|\theta \sim N\left(\frac{\theta+\tau^2 Y}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right)},\quad (B)\ \boldsymbol{\theta|Y_{mis} \sim N\left(Y_{mis},\ \tau^2\right)}.$
- **AA:** $(A')\ \boldsymbol{\tilde{Y}_{mis}|\theta \sim N\left(\frac{\tau^2(Y-\theta)}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right)},\quad (B')\ \boldsymbol{\theta|\tilde{Y}_{mis} \sim N(Y-\tilde{Y}_{mis},\ 1)}.$

An Interweaving Scheme: ASIS

- **SA:** $(A)\ Y_{mis}|\theta \sim N\left(\frac{\theta+\tau^2 Y}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad (B)\ \theta|Y_{mis} \sim N\left(Y_{mis},\ \tau^2\right).$
- **AA:** $(A')\ \tilde{Y}_{mis}|\theta \sim N\left(\frac{\tau^2(Y-\theta)}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad (B')\ \theta|\tilde{Y}_{mis} \sim N(Y-\tilde{Y}_{mis},\ 1).$
- **Let's insert (B') between (A) and (B)**

An Interweaving Scheme: ASIS

- **SA:** $(A)\ Y_{mis}|\theta \sim N\left(\frac{\theta+\tau^2 Y}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad (B)\ \theta|Y_{mis} \sim N\left(Y_{mis},\ \tau^2\right).$
- **AA:** $(A')\ \tilde{Y}_{mis}|\theta \sim N\left(\frac{\tau^2(Y-\theta)}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad (B')\ \theta|\tilde{Y}_{mis} \sim N(Y-\tilde{Y}_{mis},\ 1).$
- **Let's insert (B') between (A) and (B)**
- **Perform (A) & update $\tilde{Y}_{mis} = Y_{mis} - \theta$ use the most recent $\{Y_{mis}, \theta\}$**

$$Y_{mis}^{(t+0.5)} = \frac{\tau^2}{1+\tau^2}Y + \frac{1}{1+\tau^2}\theta^{(t)} + \sqrt{\frac{\tau^2}{1+\tau^2}}\delta_1; \quad \tilde{Y}_{mis}^{(t+1)} = Y_{mis}^{(t+0.5)} - \theta^{(t)}$$

An Interweaving Scheme: ASIS

- **SA:** $(A)\ Y_{mis}|\theta \sim N\left(\frac{\theta+\tau^2 Y}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad (B)\ \theta|Y_{mis} \sim N\left(Y_{mis},\ \tau^2\right).$
- **AA:** $(A')\ \tilde{Y}_{mis}|\theta \sim N\left(\frac{\tau^2(Y-\theta)}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad (B')\ \theta|\tilde{Y}_{mis} \sim N(Y-\tilde{Y}_{mis},\ 1).$
- Let's insert (B') between (A) and (B)
- Perform (A) & update $\tilde{Y}_{mis} = Y_{mis} - \theta$ use the most recent $\{Y_{mis}, \theta\}$

$$Y_{mis}^{(t+0.5)} = \frac{\tau^2}{1+\tau^2}Y + \frac{1}{1+\tau^2}\theta^{(t)} + \sqrt{\frac{\tau^2}{1+\tau^2}}\delta_1; \quad \tilde{Y}_{mis}^{(t+1)} = Y_{mis}^{(t+0.5)} - \theta^{(t)}$$

- Perform (B')& update $Y_{mis} = \tilde{Y}_{mis} + \theta$ use the most recent $\{\tilde{Y}_{mis}, \theta\}$

$$\theta^{(t+0.5)} = Y - \tilde{Y}_{mis}^{(t+1)} + \delta_2; \quad Y_{mis}^{(t+1)} = \tilde{Y}_{mis}^{(t+1)} + \theta^{(t+0.5)}$$

An Interweaving Scheme: ASIS

- **SA:** $(A)\ Y_{mis}|\theta \sim N\left(\frac{\theta+\tau^2 Y}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right)$, $(B)\ \theta|Y_{mis} \sim N\left(Y_{mis},\ \tau^2\right)$.
- **AA:** $(A')\ \tilde{Y}_{mis}|\theta \sim N\left(\frac{\tau^2(Y-\theta)}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right)$, $(B')\ \theta|\tilde{Y}_{mis} \sim N(Y - \tilde{Y}_{mis},\ 1)$.
- **Let's insert (B') between (A) and (B)**
- **Perform (A) & update $\tilde{Y}_{mis} = Y_{mis} - \theta$ use the most recent $\{Y_{mis}, \theta\}$**

$$Y_{mis}^{(t+0.5)} = \frac{\tau^2}{1+\tau^2}Y + \frac{1}{1+\tau^2}\theta^{(t)} + \sqrt{\frac{\tau^2}{1+\tau^2}}\delta_1; \quad \tilde{Y}_{mis}^{(t+1)} = Y_{mis}^{(t+0.5)} - \theta^{(t)}$$

- **Perform (B')& update $Y_{mis} = \tilde{Y}_{mis} + \theta$ use the most recent $\{\tilde{Y}_{mis}, \theta\}$**

$$\theta^{(t+0.5)} = Y - \tilde{Y}_{mis}^{(t+1)} + \delta_2; \quad Y_{mis}^{(t+1)} = \tilde{Y}_{mis}^{(t+1)} + \theta^{(t+0.5)}$$

- **Perform (B) to obtain the next iteration of θ**

$$\theta^{(t+1)} = Y_{mis}^{(t+1)} + \tau\delta_3$$

An Interweaving Scheme: ASIS

- **SA:** $(A)\ Y_{mis}|\theta \sim N\left(\frac{\theta+\tau^2 Y}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad (B)\ \theta|Y_{mis} \sim N\left(Y_{mis},\ \tau^2\right).$
- **AA:** $(A')\ \tilde{Y}_{mis}|\theta \sim N\left(\frac{\tau^2(Y-\theta)}{1+\tau^2},\ \frac{\tau^2}{1+\tau^2}\right), \quad (B')\ \theta|\tilde{Y}_{mis} \sim N(Y-\tilde{Y}_{mis},\ 1).$
- Let's insert (B') between (A) and (B)
- Perform (A) & update $\tilde{Y}_{mis} = Y_{mis} - \theta$ use the most recent $\{Y_{mis}, \theta\}$

$$Y_{mis}^{(t+0.5)} = \frac{\tau^2}{1+\tau^2}Y + \frac{1}{1+\tau^2}\theta^{(t)} + \sqrt{\frac{\tau^2}{1+\tau^2}}\delta_1; \quad \tilde{Y}_{mis}^{(t+1)} = Y_{mis}^{(t+0.5)} - \theta^{(t)}$$

- Perform (B')& update $Y_{mis} = \tilde{Y}_{mis} + \theta$ use the most recent $\{\tilde{Y}_{mis}, \theta\}$

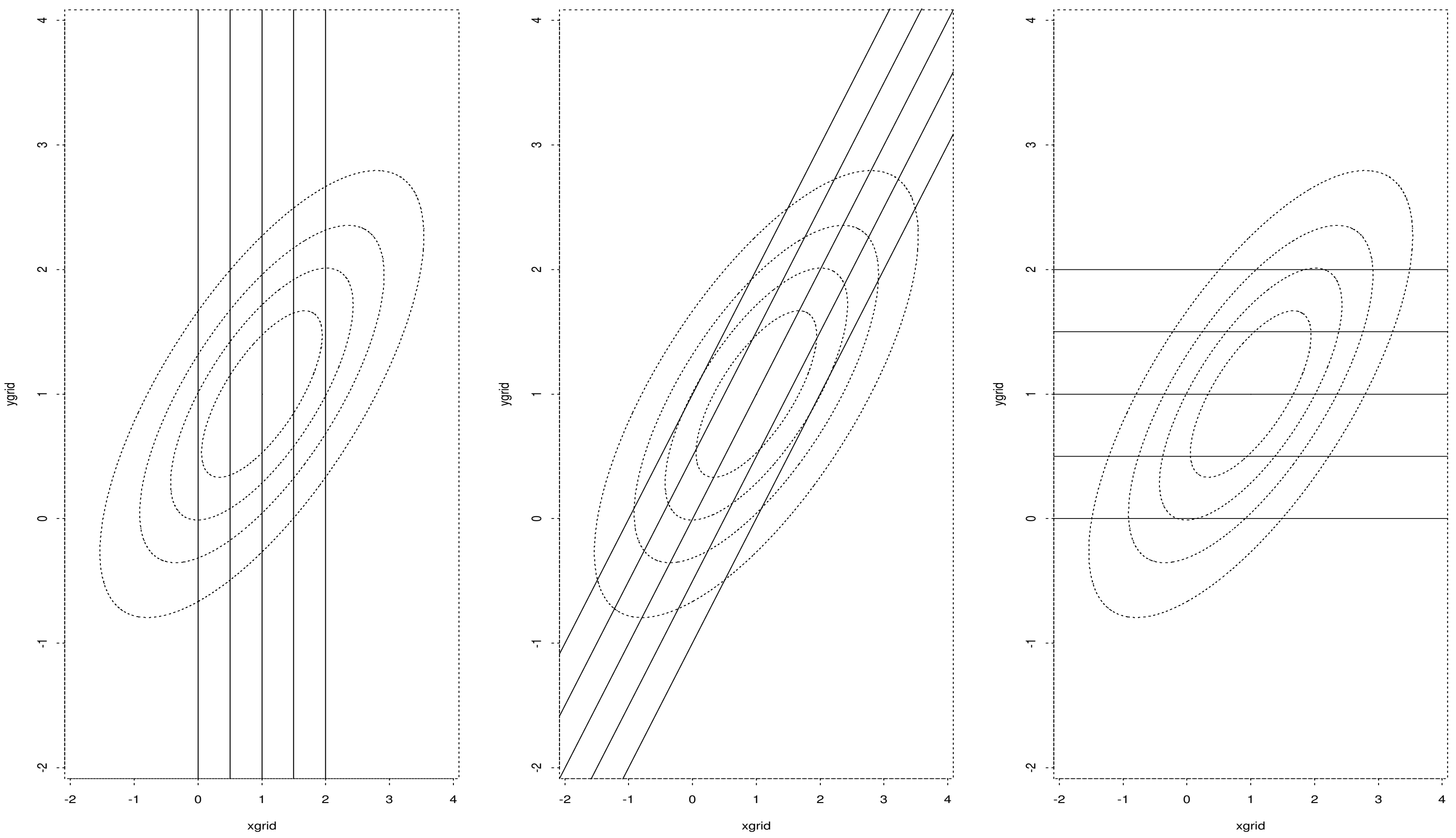
$$\theta^{(t+0.5)} = Y - \tilde{Y}_{mis}^{(t+1)} + \delta_2; \quad Y_{mis}^{(t+1)} = \tilde{Y}_{mis}^{(t+1)} + \theta^{(t+0.5)}$$

- Perform (B) to obtain the next iteration of θ

$$\theta^{(t+1)} = Y_{mis}^{(t+1)} + \tau\delta_3 = Y + \delta_2 + \tau\delta_3 \sim N(Y, 1+\tau^2).$$

- So it converges in ONE STEP: $r_{ASIS} = 0$!

Sampling Directions for the Toy Model



Sufficient and Ancillary Augmentation

Y_{mis} is sufficient for θ	$\tilde{Y}_{mis}$ is ancillary for θ
$\theta \perp Y_{obs} \mid Y_{mis}$	$\theta \perp \tilde{Y}_{mis}$ *marginally*
$\theta \rightarrow Y_{mis} \rightarrow Y_{obs}$	$\theta \longrightarrow Y_{obs}$
	$\tilde{Y}_{mis} \nearrow$
$(\theta,\ Y_{mis})$ "centered"	$(\theta,\ \tilde{Y}_{mis})$ "noncentered"
(Gelfand et al. 1995,	Papaspiliopoulos et al. 2007)

- Assume a well-defined joint distribution $p(\tilde{Y}_{mis}, Y_{mis} \mid \theta)$.
- The interwoven data augmentation scheme:

1) draw $Y_{mis} \mid (\theta,\ Y_{obs})$; 2) draw $\theta^* \mid (\tilde{Y}_{mis},\ Y_{obs})$; 3) draw $\theta \mid (Y_{mis},\ Y_{obs})$.

Alternating versus Interweaving

Alternating:

$$[Y_{mis}|\theta] \longrightarrow [\theta|Y_{mis}] \longrightarrow [\tilde{Y}_{mis}|\theta] \longrightarrow [\theta|\tilde{Y}_{mis}]. \qquad (1)$$

Alternating versus Interweaving

Alternating:

$$[Y_{mis}|\theta] \longrightarrow [\theta|Y_{mis}] \longrightarrow [\tilde{Y}_{mis}|\theta] \longrightarrow [\theta|\tilde{Y}_{mis}]. \qquad (1)$$

Interweaving (merge two middle steps):

$$[Y_{mis}|\theta] \longrightarrow [\tilde{Y}_{mis}|Y_{mis}] \longrightarrow [\theta|\tilde{Y}_{mis}].$$

Alternating versus Interweaving

Alternating:

$$[Y_{mis}|\theta] \longrightarrow [\theta|Y_{mis}] \longrightarrow [\tilde{Y}_{mis}|\theta] \longrightarrow [\theta|\tilde{Y}_{mis}]. \qquad (1)$$

Interweaving (merge two middle steps):

$$[Y_{mis}|\theta] \longrightarrow [\tilde{Y}_{mis}|Y_{mis}] \longrightarrow [\theta|\tilde{Y}_{mis}].$$

Equivalently

$$[Y_{mis}|\theta] \longrightarrow [\theta|Y_{mis}] \longrightarrow [\tilde{Y}_{mis}|Y_{mis}, \theta] \longrightarrow [\theta|\tilde{Y}_{mis}].$$

Alternating versus Interweaving

Alternating:

$$[Y_{mis}|\theta] \longrightarrow [\theta|Y_{mis}] \longrightarrow [\tilde{Y}_{mis}|\theta] \longrightarrow [\theta|\tilde{Y}_{mis}]. \qquad (1)$$

Interweaving (merge two middle steps):

$$[Y_{mis}|\theta] \longrightarrow [\tilde{Y}_{mis}|Y_{mis}] \longrightarrow [\theta|\tilde{Y}_{mis}].$$

Equivalently

$$[Y_{mis}|\theta] \longrightarrow [\theta|Y_{mis}] \longrightarrow [\tilde{Y}_{mis}|Y_{mis}, \theta] \longrightarrow [\theta|\tilde{Y}_{mis}].$$

Or

$$[\tilde{Y}_{mis}|\theta] \to [Y_{mis}|\tilde{Y}_{mis}, \theta] \to [\theta|Y_{mis}] \to [\tilde{Y}_{mis}|Y_{mis}, \theta] \to [\theta|\tilde{Y}_{mis}] \to [Y_{mis}|\theta].$$

Alternating versus Interweaving

Alternating:

$$[Y_{mis}|\theta] \longrightarrow [\theta|Y_{mis}] \longrightarrow [\tilde{Y}_{mis}|\theta] \longrightarrow [\theta|\tilde{Y}_{mis}]. \qquad (1)$$

Interweaving (merge two middle steps):

$$[Y_{mis}|\theta] \longrightarrow [\tilde{Y}_{mis}|Y_{mis}] \longrightarrow [\theta|\tilde{Y}_{mis}].$$

Equivalently

$$[Y_{mis}|\theta] \longrightarrow [\theta|Y_{mis}] \longrightarrow [\tilde{Y}_{mis}|Y_{mis}, \theta] \longrightarrow [\theta|\tilde{Y}_{mis}].$$

Or

$$[\tilde{Y}_{mis}|\theta] \rightarrow [Y_{mis}|\tilde{Y}_{mis}, \theta] \rightarrow [\theta|Y_{mis}] \rightarrow [\tilde{Y}_{mis}|Y_{mis}, \theta] \rightarrow [\theta|\tilde{Y}_{mis}] \rightarrow [Y_{mis}|\theta].$$

Essentially

$$[\tilde{Y}_{mis}|\theta] \longrightarrow [\theta|Y_{mis}] \longrightarrow [\theta|\tilde{Y}_{mis}] \longrightarrow [Y_{mis}|\theta].$$

Compare with (1)

The Benefit of "Independent Coupling" ...

Theorem 1 Given $f(Y_{obs}|\theta)$, suppose we have two augmentation schemes $Y_{mis,1}$ and $Y_{mis,2}$ with a well-defined joint distribution given (θ, Y_{obs}). Denote the geometric rate of convergence of the DA algorithm under $Y_{mis,i}$ by $r_i, \ i = 1, 2$, and the rate for the interweaving scheme by $r_{1\&2}$. Then

$$r_{1\&2} \leq R_{1,2}\sqrt{r_1 r_2},$$

where

$$R_{1,2} = \sup_{g,h} corr\{g(Y_{mis,1}), h(Y_{mis,2})\}$$

is the maximal correlation between $Y_{mis,1}$ and $Y_{mis,2}$ in their joint posterior distribution.

- Goal: Design $Y_{mis,1}$ and $Y_{mis,2}$ to be as independent as possible.
- Why Sufficiency and Ancillarity – Recall Basu's Theorem Any Complete Sufficient Statistic is independent of every Ancillary Statistic

So When/Why is $\mathcal{R}(Y_{mis}, \tilde{Y}_{mis}) = 0$?

- When

  (i) Y_{mis} is sufficient,

  (ii) $\tilde{Y}_{mis}$ is ancillary, and

  (iii) θ and Y_{mis} (same dimensions) are one-to-one given $\tilde{Y}_{mis}$, then

$$p(\tilde{Y}_{mis}, Y_{mis}|Y_{obs}) \propto p(Y_{obs}|Y_{mis})p(\tilde{Y}_{mis})p_0(\theta)J(\tilde{Y}_{mis}, Y_{mis}),$$

where $\theta = \theta(\tilde{Y}_{mis}, Y_{mis})$ is determined by $\tilde{Y}_{mis} = M(Y_{mis}; \theta)$ and

$$J(\tilde{Y}_{mis}, Y_{mis}) = \frac{\left|\det\left\{\frac{\partial M(Y_{mis};\theta)}{\partial Y_{mis}}\right\}\right|}{\left|\det\left\{\frac{\partial M(Y_{mis};\theta)}{\partial \theta}\right\}\right|}.$$

Theoretical Insights ...

- **The** a posteriori **dependence between $(\tilde{Y}_{mis}, Y_{mis})$ is determined by whether $p_0(\theta(\tilde{Y}_{mis}, Y_{mis})) \times J(\tilde{Y}_{mis}, Y_{mis})$ factors as a function of $(\tilde{Y}_{mis}, Y_{mis})$**

Theoretical Insights ...

- **The a posteriori dependence between $(\tilde{Y}_{mis}, Y_{mis})$ is determined by whether $p_0(\theta(\tilde{Y}_{mis}, Y_{mis})) \times J(\tilde{Y}_{mis}, Y_{mis})$ factors as a function of $(\tilde{Y}_{mis}, Y_{mis})$**
- **In the toy example, $p_0(\theta) \propto 1$ and $M(Y_{mis}; \theta) = Y_{mis} - \theta$, so $J(\tilde{Y}_{mis}, Y_{mis}) = 1$.**

Theoretical Insights ...

- The a posteriori dependence between $(\tilde{Y}_{mis}, Y_{mis})$ is determined by whether $p_0(\theta(\tilde{Y}_{mis}, Y_{mis})) \times J(\tilde{Y}_{mis}, Y_{mis})$ factors as a function of $(\tilde{Y}_{mis}, Y_{mis})$
- In the toy example, $p_0(\theta) \propto 1$ and $M(Y_{mis}; \theta) = Y_{mis} - \theta$, so $J(\tilde{Y}_{mis}, Y_{mis}) = 1$.
- The normality plays no role! Works when either layer of the toy model is changed from normal to any other location family.

Theoretical Insights ...

- The a posteriori dependence between $(\tilde{Y}_{mis}, Y_{mis})$ is determined by whether $p_0(\theta(\tilde{Y}_{mis}, Y_{mis})) \times J(\tilde{Y}_{mis}, Y_{mis})$ factors as a function of $(\tilde{Y}_{mis}, Y_{mis})$
- In the toy example, $p_0(\theta) \propto 1$ and $M(Y_{mis}; \theta) = Y_{mis} - \theta$, so $J(\tilde{Y}_{mis}, Y_{mis}) = 1$.
- The normality plays no role! Works when either layer of the toy model is changed from normal to any other location family.
- Intriguingly, with the Cauchy-Normal pair, the SA chain is sub-geometric (Papaspiliopoulos et. al., 2007), but ASIS still leads to iid draws. For the Normal-Cauchy pair, this is true as well except that it is the AA chain that is sub-geometric.

An important Lemma for dealing with $R_{1,2} = 1$

- **Lemma 1** Define the **maximal partial correlation (MPC)** by

$$\mathcal{R}_Y(X, Z) = \sup_{g,h} \frac{\mathrm{ECov}(g(X), h(Z)|Y)}{\sqrt{\mathrm{EV}(g(X)|Y)\mathrm{EV}(h(Z)|Y)}},$$

- Or equivalently,

$$R_Y(X, Z) = \sup_{g,h} \frac{\mathrm{Cov}(g(X) - E[g(X)|Y], h(Z) - E[h(Z)|Y])}{\sqrt{\mathrm{V}(g(X) - E[g(X)|Y])\mathrm{V}(h(Z) - E[h(Z)|Y])}}.$$

An important Lemma for dealing with $R_{1,2} = 1$

- **Lemma 1** Define the **maximal partial correlation (MPC)** by

$$\mathcal{R}_Y(X,Z) = \sup_{g,h} \frac{\mathrm{ECov}(g(X),h(Z)|Y)}{\sqrt{\mathrm{EV}(g(X)|Y)\mathrm{EV}(h(Z)|Y)}},$$

- Or equivalently,

$$R_Y(X,Z) = \sup_{g,h} \frac{\mathrm{Cov}(g(X)-E[g(X)|Y],h(Z)-E[h(Z)|Y])}{\sqrt{\mathrm{V}(g(X)-E[g(X)|Y])\mathrm{V}(h(Z)-E[h(Z)|Y])}}.$$

  Then for any (X,Y,Z) we have

$$\mathcal{R}(X,Z) \leq \mathcal{R}_Y(X,Z) + [1-\mathcal{R}_Y(X,Z)]\mathcal{R}(X,Y)\mathcal{R}(Z,Y),$$

  where the inequality is sharp in the sense that the equality holds for some non-trivial cases (e.g., when (X,Y,Z) follows a tri-variate normal with a common correlation).

A More Refined Theorem

- **More generally, given σ-algebras $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{M}$, the MPC is defined as**

$$R_{\mathcal{M}}(\mathcal{A}_1, \mathcal{A}_2) = \sup \frac{\mathrm{Cov}(X - E[X|\mathcal{M}], Z - E[Z|\mathcal{M}])}{\sqrt{\mathrm{V}(X - E[X|\mathcal{M}])\mathrm{V}(Z - E[Z|\mathcal{M}])}}$$

where sup is over all $\mathcal{A}_1$-measurable X, $\mathcal{A}_2$-measurable Z ...

- **Lemma 1' Let $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{M}$ and $\mathcal{N}$ be sub-σ-algebras on the same probability space such that $\mathcal{M} \subset \mathcal{N}$. Then**

$$\mathcal{R}_{\mathcal{M}}(\mathcal{A}_1, \mathcal{A}_2) \leq \mathcal{R}_{\mathcal{N}}(\mathcal{A}_1, \mathcal{A}_2) + [1 - \mathcal{R}_{\mathcal{N}}(\mathcal{A}_1, \mathcal{A}_2)]\mathcal{R}_{\mathcal{M}}(\mathcal{A}_1, \mathcal{N})\mathcal{R}_{\mathcal{M}}(\mathcal{A}_2, \mathcal{N}).$$

A More Refined Theorem

- More generally, given σ-algebras $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{M}$, the MPC is defined as

$$R_{\mathcal{M}}(\mathcal{A}_1, \mathcal{A}_2) = \sup \frac{\text{Cov}(X - E[X|\mathcal{M}], Z - E[Z|\mathcal{M}])}{\sqrt{\text{V}(X - E[X|\mathcal{M}])\text{V}(Z - E[Z|\mathcal{M}])}}$$

where sup is over all $\mathcal{A}_1$-measurable X, $\mathcal{A}_2$-measurable Z ...

- **Lemma 1'** Let $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{M}$ and $\mathcal{N}$ be sub-σ-algebras on the same probability space such that $\mathcal{M} \subset \mathcal{N}$. Then

$$\mathcal{R}_{\mathcal{M}}(\mathcal{A}_1, \mathcal{A}_2) \leq \mathcal{R}_{\mathcal{N}}(\mathcal{A}_1, \mathcal{A}_2) + [1 - \mathcal{R}_{\mathcal{N}}(\mathcal{A}_1, \mathcal{A}_2)]\mathcal{R}_{\mathcal{M}}(\mathcal{A}_1, \mathcal{N})\mathcal{R}_{\mathcal{M}}(\mathcal{A}_2, \mathcal{N}).$$

- **Theorem 1'** In the setting of Theorem 1, let $\mathcal{N} = \sigma(Y_{mis,1}) \cap \sigma(Y_{mis,2})$, i.e., the intersection of the σ-algebras generated by $Y_{mis,1}$ and $Y_{mis,2}$ in the joint posterior of $(\theta, Y_{mis,1}, Y_{mis,2})$. Then

$$r_{1\&2} \leq \mathcal{R}^2(\theta, \mathcal{N}) + (1 - \mathcal{R}^2(\theta, \mathcal{N}))\mathcal{R}_{\mathcal{N}}(\theta, Y_{mis,1})\mathcal{R}_{\mathcal{N}}(Y_{mis,1}, Y_{mis,2})\mathcal{R}_{\mathcal{N}}(\theta, Y_{mis,2}).$$

Component-wise Interweaving

- **Update one component of $\theta = \{\theta_1, \ldots, \theta_J\}$ at a time.**
- **Assume Y_{mis} and $\tilde{Y}_{mis,j}$ form a conditional SA and conditional AA pair for θ_j respectively, $j = 1, \ldots, J$.**

Component-wise Interweaving

- Update one component of $\theta = \{\theta_1, \ldots, \theta_J\}$ at a time.
- Assume Y_{mis} and $\tilde{Y}_{mis,j}$ form a conditional SA and conditional AA pair for θ_j respectively, $j = 1, \ldots, J$.
- Step 1. Draw $Y_{mis}|\theta^{(t)}$.

  For $j = 1, \ldots, J$, iterate the pair of steps:

  Step $(j+1)$. Draw $\theta_j^{(t+.5)} \sim P(\theta_j|\theta_{<j}^{(t+1)}, \theta_{>j}^{(t)}, Y_{mis})$.

  Step $\widetilde{(j+1)}$. Update $\tilde{Y}_{mis,j}|\theta_{<j}^{(t+1)}, \theta_j^{(t+.5)}, \theta_{>j}^{(t)}, Y_{mis}$, and then draw

$$\theta_j^{(t+1)} \sim P(\theta_j|\theta_{<j}^{(t+1)}, \theta_{>j}^{(t)}, \tilde{Y}_{mis,j});$$

  update Y_{mis} by drawing from $Y_{mis}|\theta_{\leq j}^{(t+1)}, \theta_{>j}^{(t)}, \tilde{Y}_{mis,j}$.

A General Result on Component-wise Interweaving

- **Definition:** minimal speed = 1 - maximal correlation
- $\mathcal{S}_{CIS}$: minimal speed of the component-wise interwoven algorithm

  $\mathcal{S}_j$: minimal speed for jth component defined as (under stationarity)

$$\mathcal{S}_j = 1 - \mathcal{R}_{\theta_{\neq j}}((\theta_{\neq j}, \theta_j^{(1)}), (\theta_{<j}, \theta_j^{(2)}))$$

  $\mathcal{S}_G$: minimal speed of Gibbs that iterates $p(\theta_j | \theta_{\neq j}; Y_{obs}), j = 1, \ldots, J$.

A General Result on Component-wise Interweaving

- **Definition:** minimal speed = 1 - maximal correlation
- $\mathcal{S}_{CIS}$: minimal speed of the component-wise interwoven algorithm

  $\mathcal{S}_j$: minimal speed for jth component defined as (under stationarity)

$$\mathcal{S}_j = 1 - \mathcal{R}_{\theta_{\neq j}}((\theta_{\neq j}, \theta_j^{(1)}), (\theta_{<j}, \theta_j^{(2)}))$$

  $\mathcal{S}_G$: minimal speed of Gibbs that iterates $p(\theta_j|\theta_{\neq j}; Y_{obs}), j = 1, \ldots, J.$

- Theorem 2

$$\mathcal{S}_{CIS} \geq \left(\prod_{j=1}^{J} \mathcal{S}_j\right) \tilde{\mathcal{S}}_G,$$

  where

$$\tilde{\mathcal{S}}_G = \prod_{j=1}^{J-1} \mathcal{S}_{\theta_{<j}}(\theta_{\neq j},\ \theta_{\leq j})$$

  is a lower bound on $\mathcal{S}_G$, which is sharp when $J = 2$.

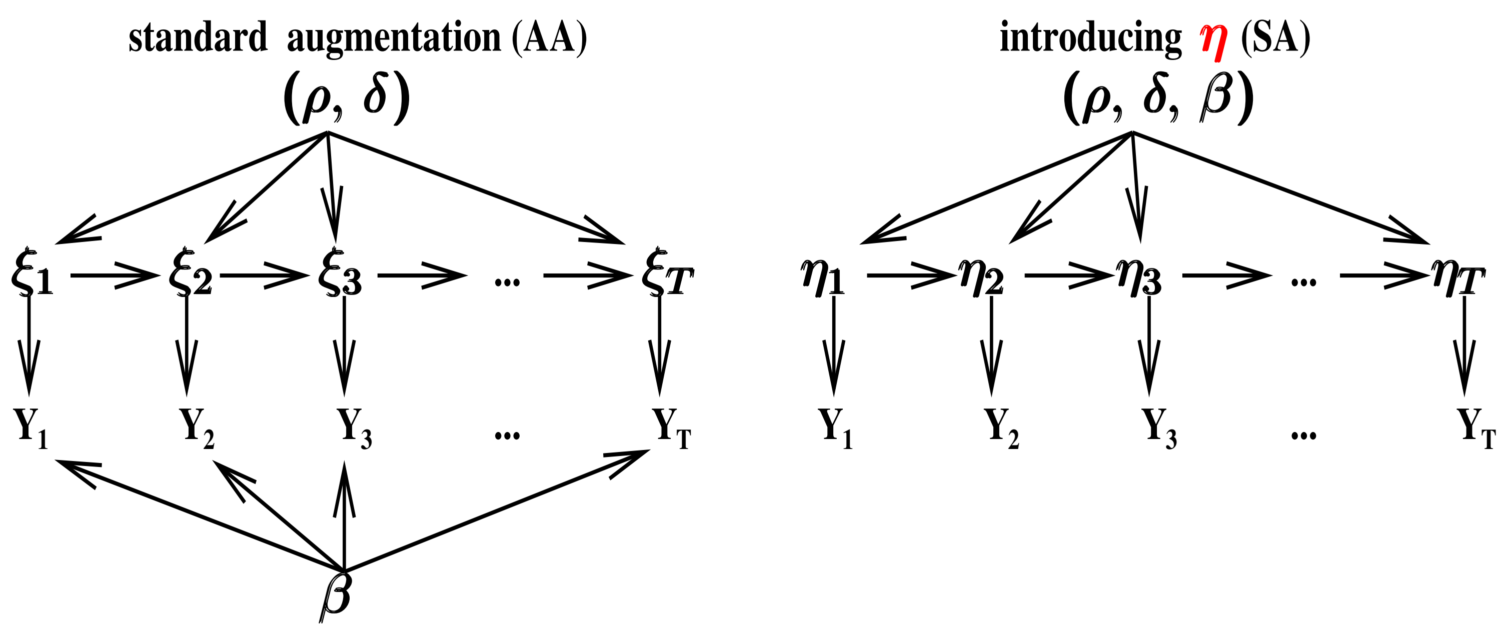
Transforming ξ: A New DA Scheme for β

$$\eta_t = \xi_t + \beta_0 + \beta_1 X_t$$

Treat $\eta = \{\eta_t\}$, instead of ξ, as the missing data:

$$Y_t \sim Pois\left(d_t e^{\eta_t}\right);$$

$$\eta_t | \eta_{<t} \sim N[\rho\eta_{t-1} + \beta_0(1-\rho) + \beta_1(X_t - \rho X_{t-1}),\ \delta^2].$$

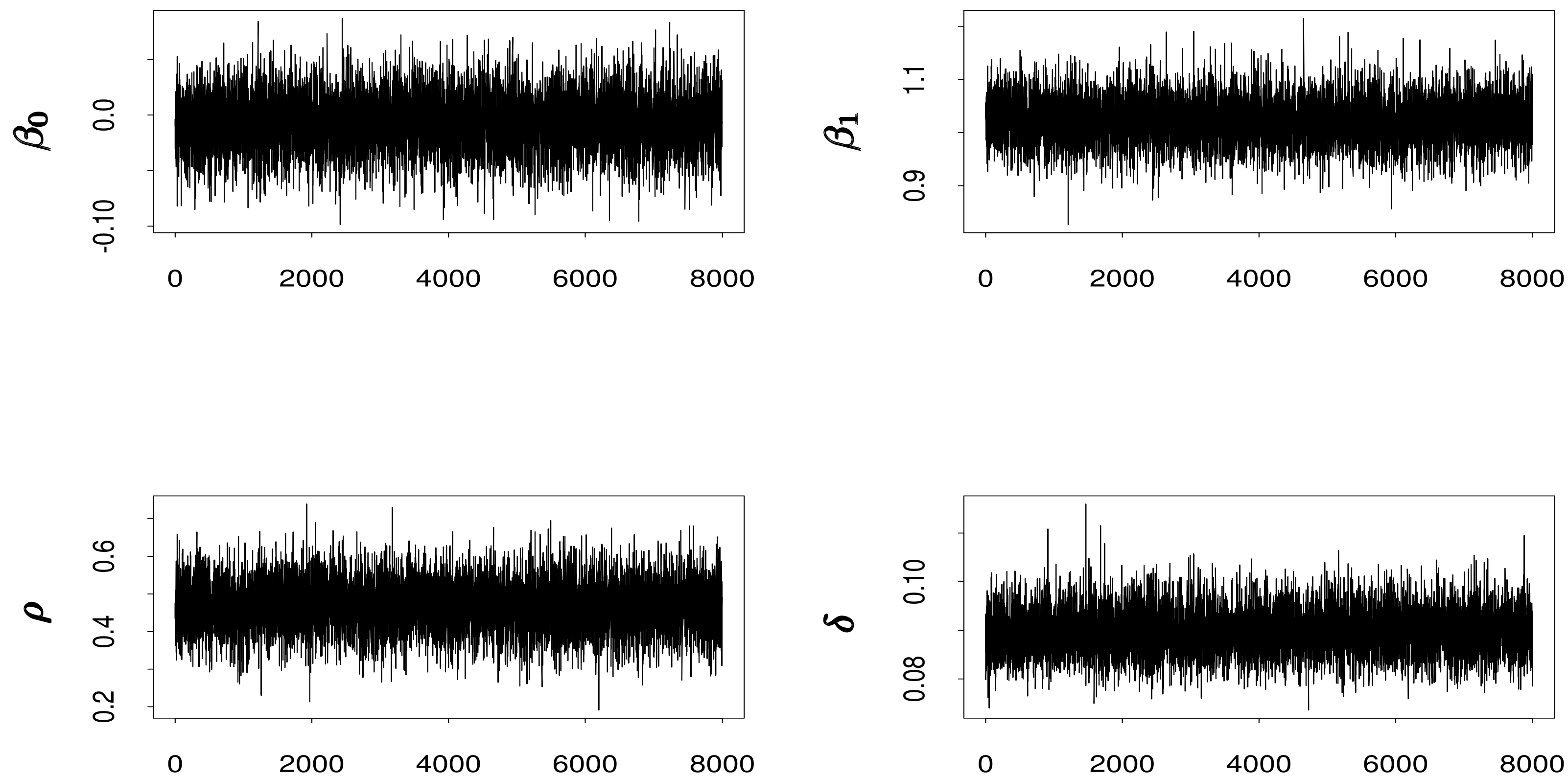


Component-wise ASIS

- $Y_t \sim Pois\left(d_t e^{\beta_0+\beta_1 X_t+\xi_t}\right); \quad \xi_t|\xi_{<t} \sim N(\rho\xi_{t-1},\ \delta^2).$
- $\eta_t = \xi_t + \beta_0 + \beta_1 X_t$
  - η is the conditional sufficient augmentation (CSA) for β;
  - ξ is the conditional ancillary augmentation (CAA) for β.

  Step 2': $\beta|(\eta,\ \rho,\ \delta)$. (In addition to Steps 1–3.)
- The Poisson likelihood does not play a role; only need linear regression.
- To keep track of ξ, set $\xi_t^{new} = \eta_t - \beta_0^{new} - \beta_1^{new} X_t$ right after Step 2'.

Performance after adding Step 2' for the same dataset, except we start from random arbitrary values:



By adding Step 2', the improvement is still dramatic after taking into account the computing time per iteration.

New Augmentation for ρ

- $Y_t \sim Pois\left(d_t e^{\beta_0+\beta_1 X_t+\xi_t}\right);\quad \xi_t|\xi_{<t} \sim N(\rho\xi_{t-1},\ \delta^2).$
- ρ has heavy autocorrelations (especially when δ is small).
- Need a new DA scheme to speed up ρ. Let

$$\zeta_t = \xi_t - \rho\xi_{t-1}.$$

- Treat $\zeta = \{\zeta_t\}$, instead of ξ, as the missing data.
  - ξ is the CSA for ρ;
  - ζ is the CAA for ρ.
- Step 3': $\rho|(\zeta,\ \beta,\ \delta)$
- This density is nonstandard; sample with an M–H step.

New Augmentation for δ

- $Y_t \sim Pois\left(d_t e^{\beta_0+\beta_1 X_t+\xi_t}\right); \quad \xi_t|\xi_{<t} \sim N(\rho\xi_{t-1},\ \delta^2).$
- Likewise design a new DA scheme for δ. Let

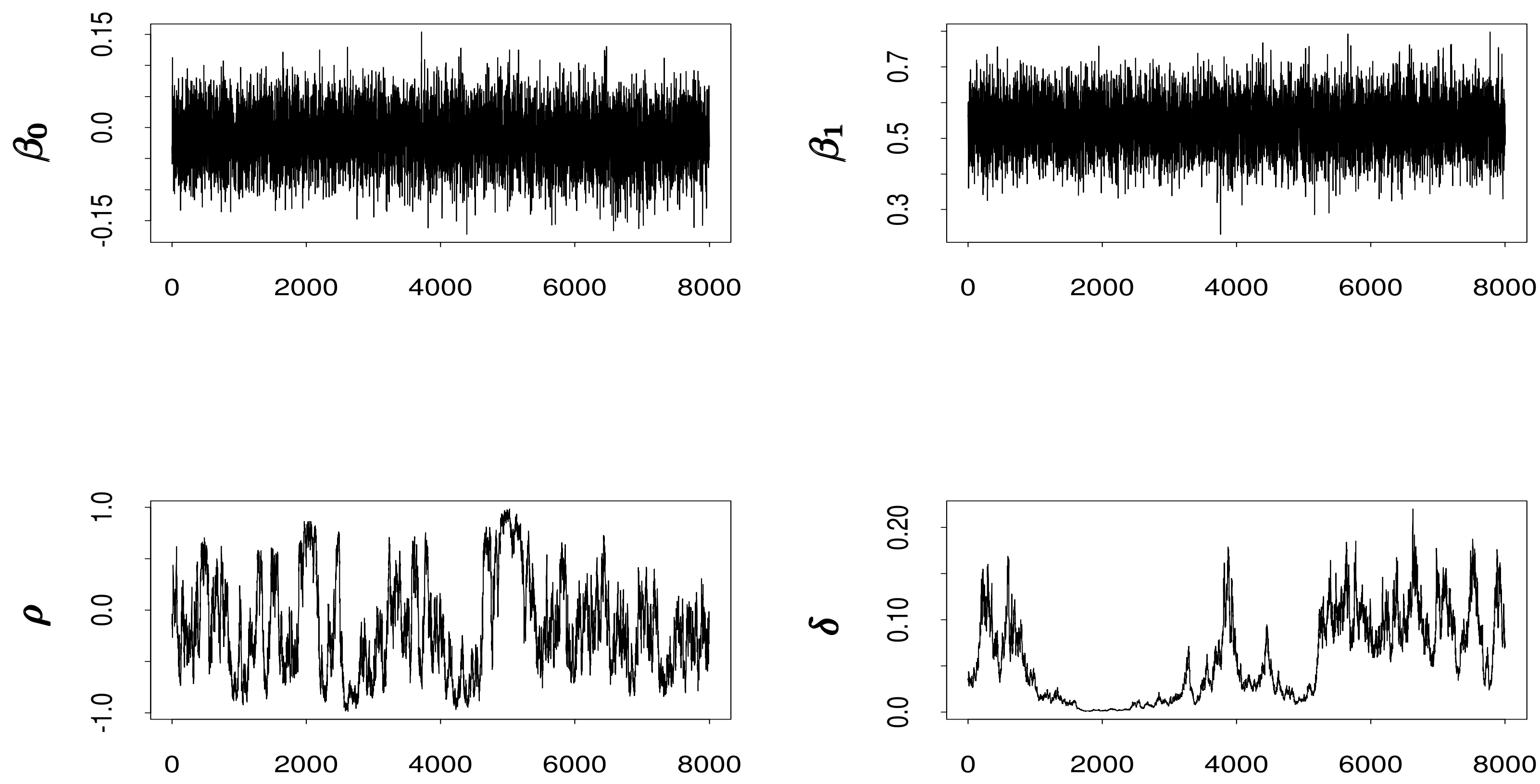
$$\kappa_t = \xi_t/\delta.$$

- Treat $\kappa = \{\kappa_t\}$, instead of ξ, as the missing data.
  - ξ is the CSA for δ;
  - κ is the CAA for δ.
- Step 3": $\delta|(\kappa,\ \beta,\ \rho)$
- Sample δ with a Metropolis step on the scale of $\log(\delta)$.

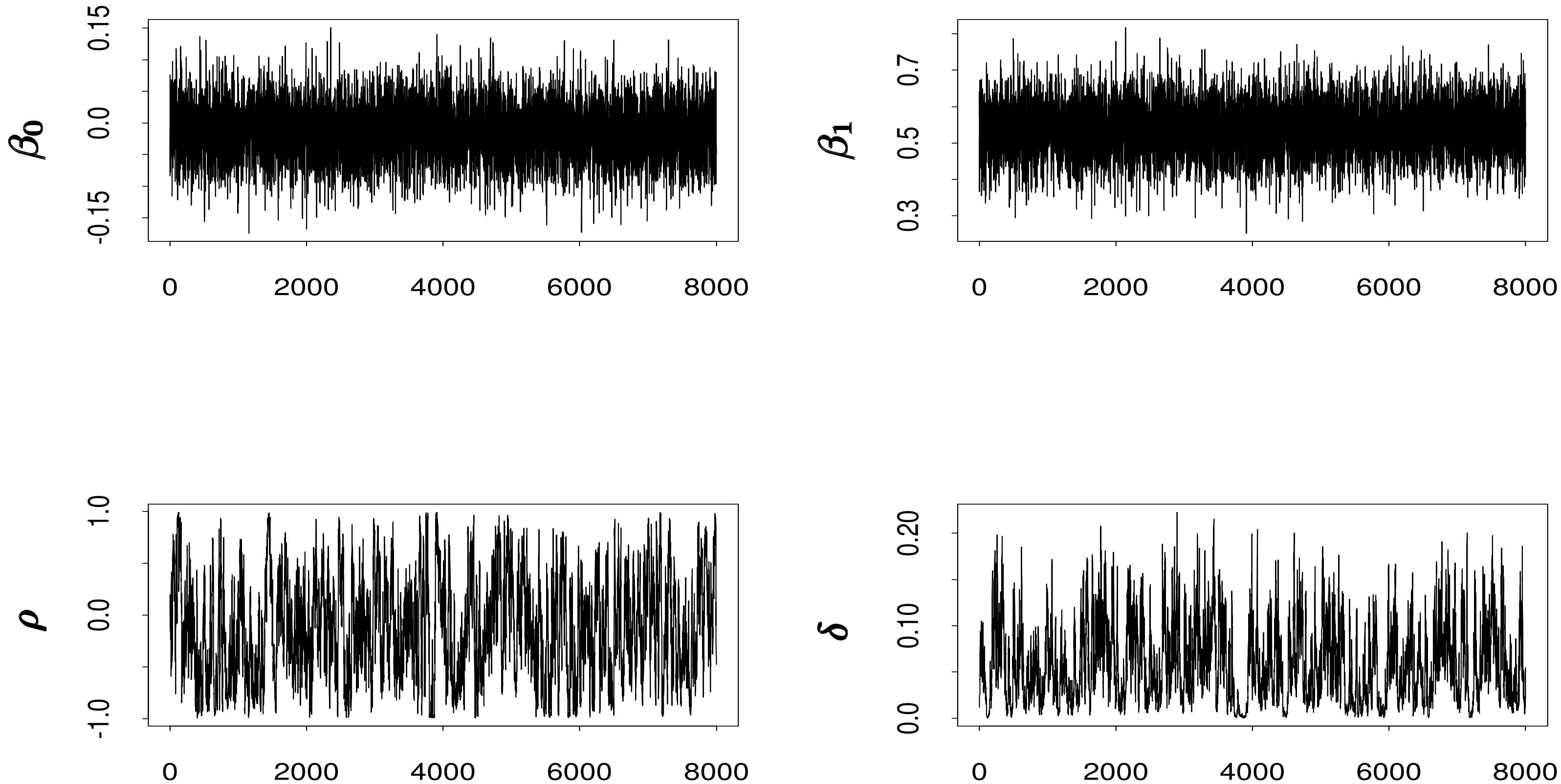
Here we generate data according to

$$T = 400, \quad d_t = 5, \quad (\beta_0, \beta_1, \rho, \delta) = (0, 0.5, 0.5, 0.01).$$

Without Step 3’ and Step 3”



With Step 3' and Step 3"



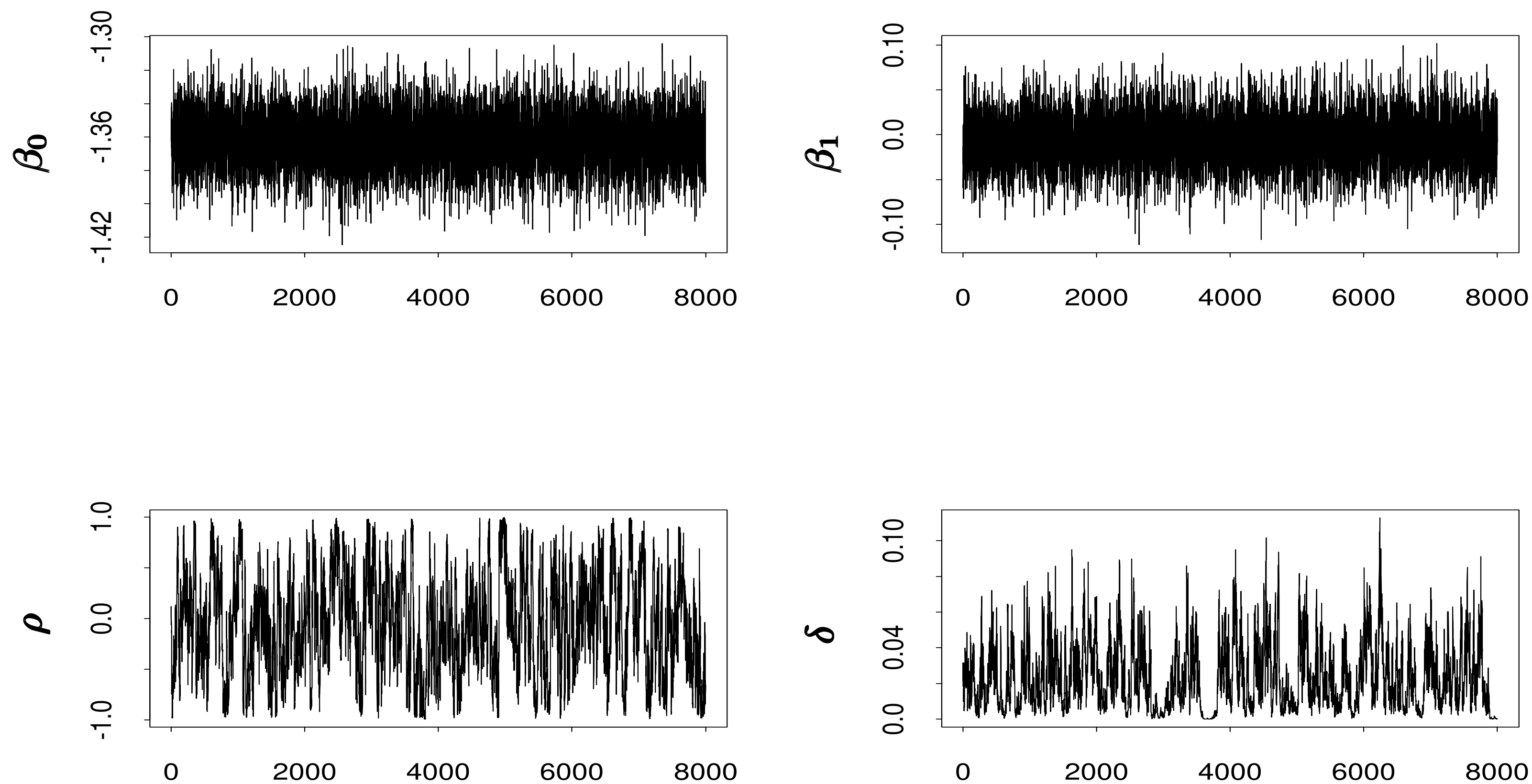
Final Algorithm

- **Step 1**: draw $\xi|(\beta,\ \rho,\ \delta)$.
- **Steps 2–2' and 3–3"**: draw

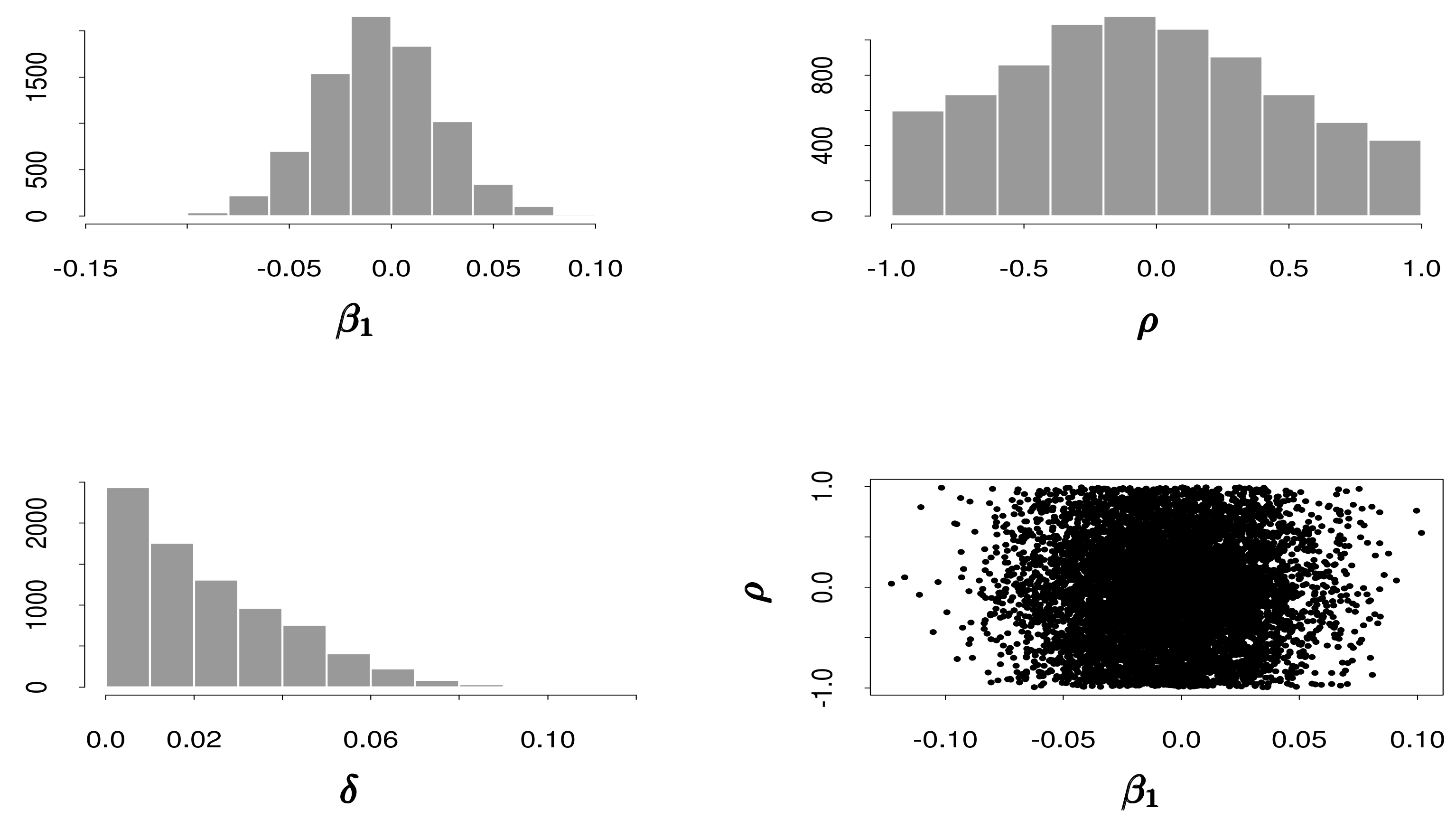
	CSA	CAA
$\beta|$	$\eta:\ \eta_t = \xi_t + \beta_0 + \beta_1 X_t$	ξ *
$\rho|$	ξ *	$\zeta:\ \zeta_t = \xi_t - \rho\xi_{t-1}$
$\delta|$	ξ *	$\kappa:\ \kappa_t = \xi_t/\delta$

  - *: Steps performed by the standard Gibbs sampler.
  - $\rho|(\xi,\ \beta,\ \delta)$ and $\delta|(\xi,\ \beta,\ \rho)$ are combined into $(\rho,\ \delta)|(\xi,\ \beta)$.

The Neutron Star/Quark Star Candidate



Posterior Summary: Intensity Does Not Change



Working Parameter: A Quick Review

PX–DA: the parameter-expanded DA (Liu and Wu 1999, Meng and van Dyk, 1999).

- **Original model** $(\theta,\ Y_{mis},\ Y_{obs})$.
- **Expanded model** $(\theta,\ \alpha,\ \tilde{Y}_{mis} = M_\alpha(Y_{mis}),\ Y_{obs})$.
  - **Key: the observed data model** $(\theta,\ Y_{obs})$ **should be preserved.**
- **PX–DA iterates between two steps:**
  1. **draw** $(\alpha,\ \tilde{Y}_{mis})|(\theta,\ Y_{obs})$;
  2. **draw** $(\alpha,\ \theta)|(\tilde{Y}_{mis},\ Y_{obs})$.
- – **The user may specify a prior** $p(\alpha)$ **for the expansion parameter** α.
  - **"Haar measure is the best" (Liu and Wu 1999).**

Optimality: Sometimes ASIS= Optimal PX–DA

Interweaving SA and AA not only is robust, but also optimal sometimes.

Theorem 3 **In Theorem 1, assume in addition that the DA schemes are linked by a 1–1 mapping $\tilde{Y}_{mis} = M_\theta(Y_{mis})$, and**

1. **M_θ is a locally compact group with a unimodular Haar measure; and**
2. **the prior $p(\theta)$ w.r.t. the Haar measure satisfies $p(\theta \cdot \theta') \propto p(\theta)p(\theta')$.**

Then algorithm ASIS coincides with the optimal PX–DA algorithm (i.e., PX–DA with the Haar measure prior) for the expanded model $(\theta,\ \alpha,\ \tilde{Y}_{mis},\ Y_{obs})$, where α is the expansion parameter and $\tilde{Y}_{mis} = M_\alpha(Y_{mis})$.

- **In particular $r_{ASIS} \leq \min\{r_{SA},\ r_{AA}\}$.**
- **Condition 2 is satisfied by $p(\theta) \propto 1$, for example.**

To Center, or Not To Center?

To Center, or Not To Center?

USE BOTH!

Conclusion

To Center, or Not To Center?

USE BOTH!

Try it – if it does not work, I will refund the time you listen to this talk!