# From Fields to Trees

Firas Hamze, Nando De Freitas
UAI 2004

Presented by: samy

February 20, 2014

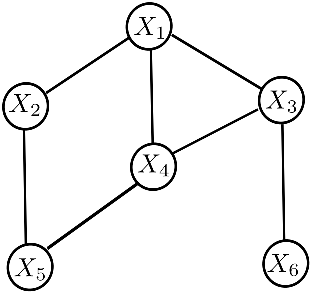
# Papers

From Fields to Trees
Firas, Hamze 2004

Covariance Structure of the Gibbs Sampler ...
Liu, Wong, Kong 1994

Markov Chain Strategies in Scientific Computing
Jun Liu, 2001

# The Problem




Given: Graphical Model with known Parameters.

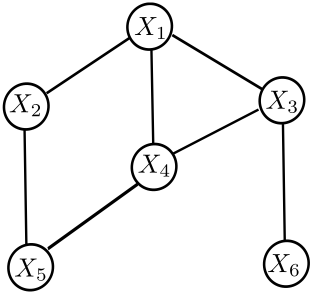
**Goal:** Estimate $\mathbb{E}[h(X)]$
E.g.: $\mathbb{1}_A(x)$, $\sum_i X_i$

Generate Samples: $X^{(1)}, X^{(2)}, \ldots X^{(N)}$

Empirical Estimator

$$\mu_0 = \frac{1}{N} \sum_{i=1}^{N} h(X^{(i)})$$

# Gibbs Sampling




Current Sample: $X^{(t)}$

$X_1^{(t+1)} \mid X_{2,3,4,5,6}^{(t)}$

$X_2^{(t+1)} \mid X_1^{(t+1)}, X_{3,4,5,6}^{(t)}$

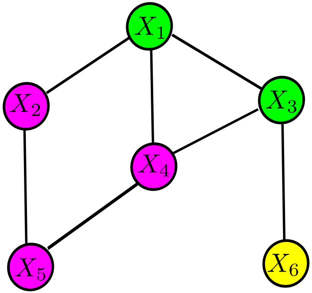
$\dots$

$X_6^{(t+1)} \mid X_{1,2,3,4,5}^{(t+1)}$

Next Sample: $X^{(t+1)}$

$X^{(1)} \to X^{(2)} \to \dots$ is a Markov Chain with eqlb$^{\mathrm{m}}$ distribution $P$.

# Blocked Gibbs Sampling




Current Sample: $X^{(t)}$

$X_{1,3}^{(t+1)} \mid X_{2,4,5,6}^{(t)}$

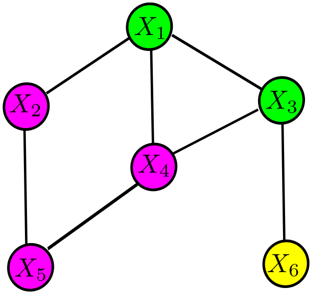
$X_{2,4,5}^{(t+1)} \mid X_{1,3}^{(t+1)}, X_6^{(t)}$

$X_6^{(t+1)} \mid X_{1,2,3,4,5}^{(t+1)}$

Next Sample: $X^{(t+1)}$

$X^{(1)} \to X^{(2)} \to \ldots$ is a Markov Chain with eqlb$^{\text{m}}$ distribution $P$.

# Blocked Gibbs Sampling

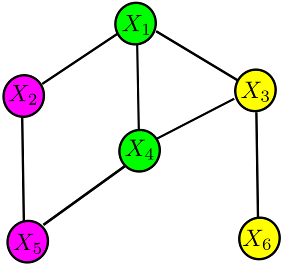


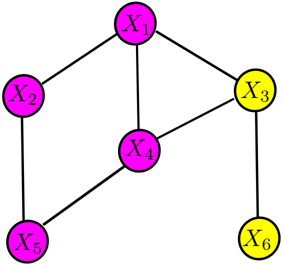

Why Blocked Gibbs ?

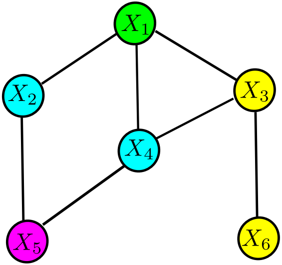
Sampling from conditional distribution is difficult.

BUT, Chain mixes faster $\implies$ better samples.

# How to Block (Partition) the Variables?



# Outline

1. Digression
   - MCMC Theory
   - Rao-Blackwellisation
2. Primary Idea
   - Tree Partitions
3. Hamze & Freitas, 2004
4. My 701 Course Project

# MCMC Theory (Jun Liu, 2001)

$$X^{(1)} \to X^{(2)} \to \dots \to X^{(t)} \to X^{(t+1)} \to \dots$$

$$\pi \leftarrow \text{Equilibrium Distribution}$$

# MCMC Theory (Jun Liu, 2001)

$$X^{(1)} \to X^{(2)} \to \cdots \to X^{(t)} \to X^{(t+1)} \to \ldots$$

$$\pi \leftarrow \text{Equilibrium Distribution}$$

$L_0(\pi) = \{h : \Omega \to \mathbb{R} : \ \ \mathbb{E}_\pi h(X) = 0, \ \ \mathbb{V}_\pi h(X) < \infty\}$
$\langle h, g\rangle = \text{Covar}_\pi(h(X), g(X))$
$\big(L_0(\pi), \langle\cdot,\cdot\rangle\big)$ is a Hilbert Space.

# MCMC Theory (Jun Liu, 2001)

$$X^{(1)} \to X^{(2)} \to \cdots \to X^{(t)} \to X^{(t+1)} \to \dots$$

$\pi \leftarrow$ Equilibrium Distribution

$$L_0(\pi) = \{h : \Omega \to \mathbb{R} : \ \ \mathbb{E}_\pi h(X) = 0, \ \ \mathbb{V}_\pi h(X) < \infty\}$$

$$\langle h, g \rangle = \mathrm{Covar}_\pi(h(X), g(X))$$

$\big(L_0(\pi), \langle \cdot, \cdot \rangle\big)$ is a Hilbert Space.

$$\| h \|^2 = \langle h, h \rangle = \mathrm{Covar}_\pi(h(X), h(X)) = \mathbb{V}_\pi(h(X))$$

# Forward Operator

Define $F : L(\pi) \to L(\pi)$, $\quad [Fh](z) = \mathbb{E}[h(X^{(1)})|X^{(0)} = z]$

# Forward Operator

Define $F : L(\pi) \to L(\pi), \quad [Fh](z) = \mathbb{E}[h(X^{(1)})|X^{(0)} = z]$

Norm of operator: $\|F\| = \sup_{\|h\| \leq 1} \|Fh\|$.

# Forward Operator

Define $F : L(\pi) \to L(\pi)$, $\quad [Fh](z) = \mathbb{E}[h(X^{(1)})|X^{(0)} = z]$

Norm of operator: $\|F\| = \sup_{\|h\| \leq 1} \|Fh\|$.

$$\begin{aligned} \|Fh\|^2 &= \mathbb{V}_{X^{(0)} \sim \pi}(\mathbb{E}[t(X^{(1)})|X^{(0)}]) \\ &\leq \mathbb{V}_{X^{(0)} \sim \pi}(h(X^{(1)})) = \mathbb{V}_{X^{(1)} \sim \pi}(h(X^{(1)})) = \|h\|^2 \end{aligned}$$

Hence, $\|F\| \leq 1$.

# Forward Operator

Define $F : L(\pi) \to L(\pi)$, $\quad [Fh](z) = \mathbb{E}[h(X^{(1)})|X^{(0)} = z]$

Norm of operator: $\|F\| = \sup_{\|h\| \leq 1} \|Fh\|$.

$$\begin{aligned}\|Fh\|^2 &= \mathbb{V}_{X^{(0)} \sim \pi}(\mathbb{E}[t(X^{(1)})|X^{(0)}]) \\ &\leq \mathbb{V}_{X^{(0)} \sim \pi}(h(X^{(1)})) = \mathbb{V}_{X^{(1)} \sim \pi}(h(X^{(1)})) = \|h\|^2\end{aligned}$$

Hence, $\|F\| \leq 1$.

If MC is reversible, $F$ is self-adjoint $\implies \|F^n\| = \|F\|^n$.

# Convergence of a MC

**Rate of Convergence:** $\|F\|$

**Theorem:** Let $X^{(0)} \sim P^{(0)}$. $P^{(n)}, \mathbb{E}^{(n)}$ : n$^{\text{th}}$ step evolution.

$$|\mathbb{E}^{(n)} h(X) - \mathbb{E}_\pi h(X)| \leq C \|F\|^n \|h\|.$$

# Convergence of a MC

**Rate of Convergence:** $\|F\|$

**Theorem:** Let $X^{(0)} \sim P^{(0)}$. $P^{(n)}, \mathbb{E}^{(n)}$ : n$^{th}$ step evolution.

$$|\mathbb{E}^{(n)} h(X) - \mathbb{E}_\pi h(X)| \leq C \|F\|^n \|h\|.$$

$$\|F\| = \sup_{f,g} \text{Corr}(f(X^{(t+1)}), g(X^{(t)}))$$

# Convergence of a MC

**Rate of Convergence:** $\|F\|$

**Theorem:** Let $X^{(0)} \sim P^{(0)}$. $P^{(n)}, \mathbb{E}^{(n)}$ : n$^{\text{th}}$ step evolution.

$$|\mathbb{E}^{(n)} h(X) - \mathbb{E}_\pi h(X)| \le C \|F\|^n \|h\|.$$

$\|F\| = \sup_{f,g} \text{Corr}(f(X^{(t+1)}), g(X^{(t)}))$

Finite State space: $\|F\| \equiv 2^{\text{nd}}$ eigval of Transition Matrix.

# Rao-Blackwellisation

$(X, Y) \sim P$. Samples:
$\{(X^{(1)}, Y^{(1)}), (X^{(2)}, Y^{(2)}), \ldots (X^{(N)}, Y^{(N)})\}$

**Goal:** Estimate $\mathbb{E}_P[h(X)]$

Empirical Estimator $\quad \mu_0 = \frac{1}{N} \sum_{i=1}^{N} h(X^{(i)})$

Rao-Blackwellised Estimator $\quad \mu_{rb} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[h(X) | Y^{(i)}\right]$

# Rao-Blackwellisation

$(X, Y) \sim P$. Samples:
$\{(X^{(1)}, Y^{(1)}), (X^{(2)}, Y^{(2)}), \ldots (X^{(N)}, Y^{(N)})\}$

Independent samples: RB Estimator is better.

# Rao-Blackwellisation

$(X, Y) \sim P$. Samples: $\{(X^{(1)}, Y^{(1)}), (X^{(2)}, Y^{(2)}), \dots (X^{(N)}, Y^{(N)})\}$

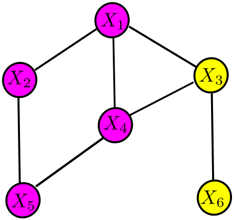
Independent samples: RB Estimator is better.

2-Stage Gibbs: RB Estimator is better (LKW 1994).

$$\mathbb{V}(\mu_0) - \mathbb{V}(\mu_{rb}) = \frac{\mathbb{V}(h(X))}{n^2} \Big( n(1 - \|F\|) + 2(n-1)(\|F\| - \|F\|^2) \\ \dots + (\|F\|^{n-1} - \|F\|^n) \Big)$$

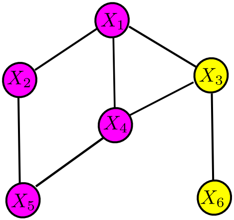
# Outline

# Primary Idea




Correlated Partitions $\implies$
Successive Samples Correlated $\implies$
Large $\|F\|$

# Primary Idea

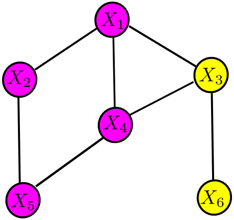



Correlated Partitions $\implies$
Successive Samples Correlated $\implies$
Large $\|F\|$

Blocked Gibbs : $\|F\|$ is determined by the partitions.

Partition the Graph so that $\|F\|$ is small !

# Primary Idea



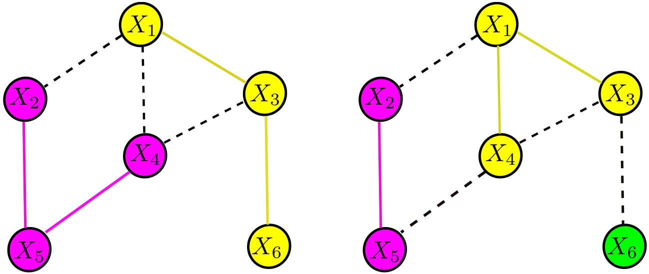

Correlated Partitions $\implies$
Successive Samples Correlated $\implies$
Large $\|F\|$

Blocked Gibbs : $\|F\|$ is determined by the partitions.

Partition the Graph so that $\|F\|$ is small !

Is this the best strategy ?

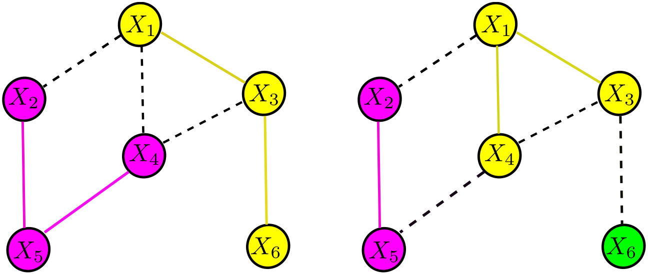
# Tree Partitions




Lets focus on Tree Partitions.

Computing conditional distribution on trees is easy.

# Tree Partitions

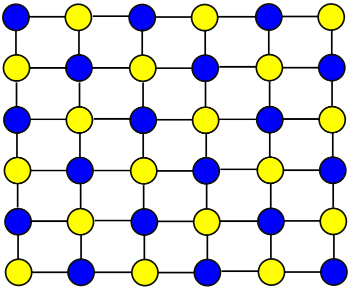



Lets focus on Tree Partitions.

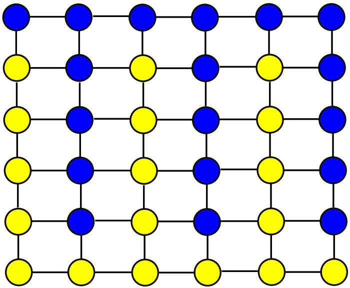
Computing conditional distribution on trees is easy.

Best Tree Partition ? Not easy.

# Grid Graph

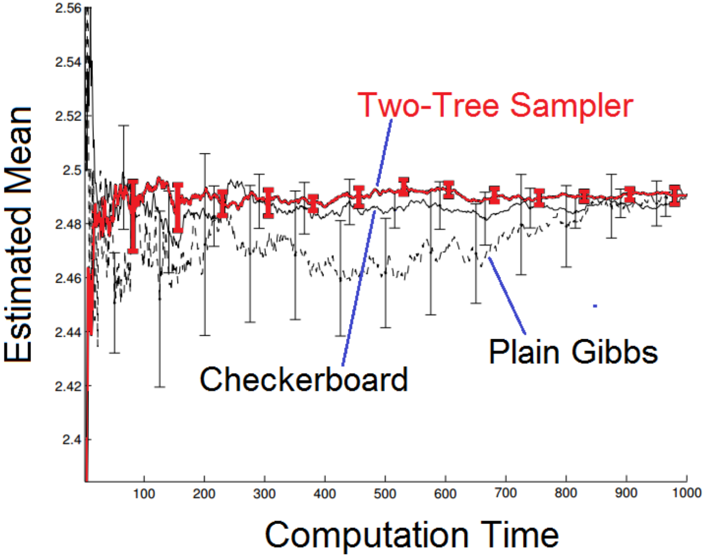


Checkerboard



Two-Trees

# Results




H/deF use a a Rao-Blackwellised Estimator.

# Outline

Incorporate Correlations via Weights.