Efficient Sampling of Inverse Correlation Matrices and its Applications in Bayesian Modeling of Gene Interaction in Early Phase Genomic Experiments

Nitai D. Mukhopadhyay

Virginia Commonwealth University

Co-author : Sarat C. Dass

Michigan State University

Presentation at : Biopharmaceutical Applied Statistics Symposium XIV.

Nov 2009

Outline

- Explanation of the title :
  - Gene interaction studies.
  - Why inverse correlation matrix.
  - Bayesian Analysis and need for a sampling scheme.
- Previous literature
  - Meng et al
  - Wong et al
  - Limitations and challenges.
- Proposed Algorithm
  - Reversible Jump Markov Chain.
  - Cholesky Decomposition of inverse correlation.
  - Inverse correlation selection prior.
  - Sampling Algorithm.
- Application to Gene interaction experiment.
  - Androgen pathway

Gene Interaction/pathway studies :

Evolution of the Androgen Receptor Pathway during Progression of Prostate Cancer. Peter J.M. Hendriksen et al , Cancer Res 2006; 66: (10). May 15, 2006

- Involvement of Androgen receptor pathway in initiation and progression of prostate cancer.
- 200 genes that are androgen responsive in prostate cancer cell lines and/or xenograft.
- Hierarchical clustering, up-regulation and down-regulation in a gene-wise analysis was determined based on ratio of expression between cancer cells and normal cells above 1.62 (2 log 0.7).

Conclusion

- Specific sets of androgen receptor pathway genes are down-regulated during the progression from well-differentiated prostate carcinoma to high grade prostate carcinoma.

Pathway studies cont.

Genome-wide expression profiling reveals transcriptomic variation and perturbed gene networks in androgen-dependent and androgen-independent prostate cancer cells : A.P. Singh et al. / Cancer Letters 259 (2008) 28–38

- 35 cancer cases, hybridized to the HGU133 plus2 gene chips (Affymetrix)
- Paired t-test between benign and prostate cancer cells was used to determine significant difference.
- Pathway prediction analysis using 'Ingenuity Pathway Analysis'.

Conclusion

- Multiple genes were identified as differentially expressed including many tumor suppressor genes.
- Pathway prediction analysis identified several signaling pathways to be perturbed.

"The Functional Analysis of a network identified the biological functions and/or diseases that were most significant to the molecules in the network. The network molecules associated with biological functions and/or diseases in Ingenuity's Knowledge Base were considered for the analysis. Right-tailed Fisher's exact test was used to calculate a p-value determining the probability that each biological function and/or disease assigned to that network is due to chance alone." –Pathway Assist website.

Towards a more comprehensive model :

- Need to account for dependence among the genes in order to make conclusion about pathways.
- The model will be high dimensional as gene interaction is modeled through their covariance matrix.
- Bayesian approach would need a prior on the covariance parameters and posterior sampling.

Previous work on covariance selection models

Covariance and correlation matrices play an important role in statistical inference. Modeling covariance and correlation matrices is a difficult task due to the non-negative definiteness constraint placed on these matrices.

1. Barnard, McCullogh and Meng (2000)

- Barnard et al. (2000) modeled the covariance matrix by decomposing into the variance components and the correlation matrix, R.
- Prior on R are developed so that there is shrinkage towards 0 for each entry.

Drawbacks:

- Elements of R are updated one at a time.
- To preserve non-negative definiteness, each entry is constrained to lie in an interval.
- This interval has to be computed at each update resulting in a slow updating algorithm.

Previous work on covariance selection models (cont.)

2. Wong, Carter and Kohn (2003)

- Wong, Carter and Kohn (2003) developed statistical inference for the *inverse* covariance matrix, W, in a graphical network.
- Prior elicited in terms of zero and non-zero entries of W.
- Again, non-negative definiteness of W constrain the entries of W to belong to an interval.
- Elements of W are updated one at a time. The updating algorithm is slow to converge.

3. Pitt, Chan and Kohn (2006)

- Pitt, Chan and Kohn (2006) focuses on inference on the inverse correlation matrix, W, such that $W^{-1} = R$.
- More complicated since diagonal entries of W^{-1} must be unity.
- Their updating algorithm is also slow to converge.

Why inverse correlation matrix :

Let $X = (X_1, X_2, \ldots, X_p)^T$ denote a p-variate r.v. in R^p .

$$\text{Density} \quad : \phi_p(x_1, x_2, \ldots, x_p \mid R) = \frac{1}{(2\pi)^{p/2} \det(R)^{1/2}} \exp\{-\frac{1}{2} x^T R^{-1} x\}$$

$$R = ((r_{ij})) : \text{Symmetric p d Correlation matrix.}$$

$$W = ((w_{ij})) = R^{-1}$$

then $w_{ij} = 0 \Leftrightarrow X_i$ and X_j are conditionally independent given the rest of the X_ks, $k \neq \{i, j\}$.

Bayesian Inference :

- Likelihood : Sampling distribution of data $\varphi(X|R)$.
- Prior for R : $\pi(R)$.
- Posterior : $p(R \mid X) \propto \pi(R)\phi(X \mid R)$
- Posterior integration problem : Almost all inference requires computation of the integral :

$$\int f(R)p(R \mid X)dR$$

Typical problems :

- Analytically intractable integral.
- f an p are generally nice function.
- R is very high dimensional.

Posterior MCMC

- Monte Carlo integration :

$$\int f(R)p(R \mid X)dR \approx \frac{1}{M}\sum_{i=1}^{M} f(R_i) \text{ where } R_i \sim p(R \mid X)$$

- How to generate

$$R_i \sim p(R \mid X)$$

  - Construct a Markov chain with the stationary distribution $p(R \mid X)$ and generate $R_i, i = 1,2,...$

Model selection problem :

- For the graphical model problem on p nodes, the correlation matrix has p(p-1)/2 parameters.
- Natural networks (gene pathways) are often sparse.

Suppose J_i denote the set of all correlation matrices subject to a specific configuration of the inverse correlation matrix consisting of zeros and nonzero values. Our models are :

$$M_i : \quad R \in J_i \text{ for all possible } J_i \text{ , } i = 1,...,M$$

- Models are of different dimension.
- Parameter space is $I \otimes R$, I is the model indicator, R is the correlation matrix subject to configuration J_i.

Goal :

- To develop a sampling scheme that is able to :
  - Sample within the restricted domain of R or W matrix.
  - Update multiple entries of the matrix at a time.
  - Jump between all sparsity configurations of W.

Reversible Jump MCMC :

- Green (1995) and Green and Richardson (1997) developed the RJMCMC approach for Bayesian inference.
- Let x and y be elements of the model space (with possibly differing dimensions).
- The RJMCMC approach proposes a move, say m, with probability r_m.
- The move m takes x to y via the proposal distribution $q_m(x, y)$.
- Time reversibility requires accepting y with probability

$$\alpha(x, y) = \min\{\, 1, \frac{\pi(y \mid data\ )}{\pi(x \mid data\ )} \frac{r_{m'} q_{m'}(y, x)}{r_m q_m(x, y)}$$

$$\pi(x \mid data\ ) : \text{posterior of } x \text{ given data.}$$

$$q_{m'}(y,x) = \text{probabilit y of moving from y to x.}$$

RJMCMC continued :

- Moves *m* and m_0 must be reversible: $q_m(x, y)$ and $q_{m'}(y, x)$ must have the same support.
- If *y* represents the higher dimensional model, we can first sample *u* from a proposal $q_0(x, u)$ then obtain *y* as a one-to-one function of *(x, u).*
- In that case,

$$q_m(x, y) = q_0(x, u) / \det\left(\frac{\partial y}{\partial (x, u)}\right)$$

Cholesky decomposition of Inverse Correlation matrix

(Mukhopadhyay, Dass (2009) Tech report MSU.)

$R^{-1} = W = LL'$: Cholesky decomposition.

Denote :

$$W = \begin{bmatrix} * & * & * \\ * & w_{jj} & w_j' \\ * & w_j & W_{jj} \end{bmatrix} \quad and \quad L = \begin{bmatrix} * & 0 & 0 \\ * & l_{jj} & 0 \\ * & l_j & L_{jj} \end{bmatrix}$$

Then :

$$l_{jj}^2 = 1 + l_j' (L_{jj} L_{jj}')^{-1} l_j$$

- This allows us to treat the l_{ij} s for $i > j$ as free (i.e., unconstrained) parameters with each l_{ij} taking values on R.
- The elements of l_{ij} involve $l_{kk'}$s for indices $k > k' > j$ only.

Inverse correlation selection prior

(Mukhopadhyay, Dass (2009) Tech report MSU.) :

Let J denote a configuration of zeros and nonzeros in the W matrix. All the models are indexed by it's configuration.

Prior on R :

$$g(R \mid J) \propto \prod_{j=1}^{p-1} (\det(R_{[I_j, I_j]}))^{-1/2}$$

Where $R_{[I_j, I_j]}$ is the submatrix of R_{jj} with rows and columns correspond ing to the non - zero elements of L.

$$\pi_0(J \mid N(J) = h) = \binom{H}{h}^{-1}, \text{ and}$$

$$\pi_0(N(J) = h \mid \psi) = \binom{H}{h} \psi^h (1-\psi)^{H-h}, \text{ and}$$

$$\pi_0(\psi) = Uniform \quad (0,1)$$

Prior cont.

- Prior on R is proper and the normalizing constant can be computed in a close form :

$$V(J) = \int g(R \mid J) dr_J = \prod_{j=1}^{p-1} Beta\left(\frac{n_j}{2}, 1 + \frac{n_j}{2}\right) \frac{2\pi^{n_j/2}}{n_j \Gamma(n_j/2)}$$

where n_j = no. of nonzero entries in j - th column of L.

Reversible Jump Sampling Algorithm

- **Remove Zero:**

  In this updating step, (J,W) $\rightarrow$(J*,W*) by increasing the number of nonzero entries j-th column of L.

- **Add Zero:**

  In this updating step, (J,W) $\rightarrow$(J*,W*) by increasing the number of zero entries j-th column of L.

- **Zero Unchanged:**

  In this updating step, (J,W) $\rightarrow$(J*,W*) by updating the nonzero entries j-th column of L without changing the zero pattern.

Algorithm :

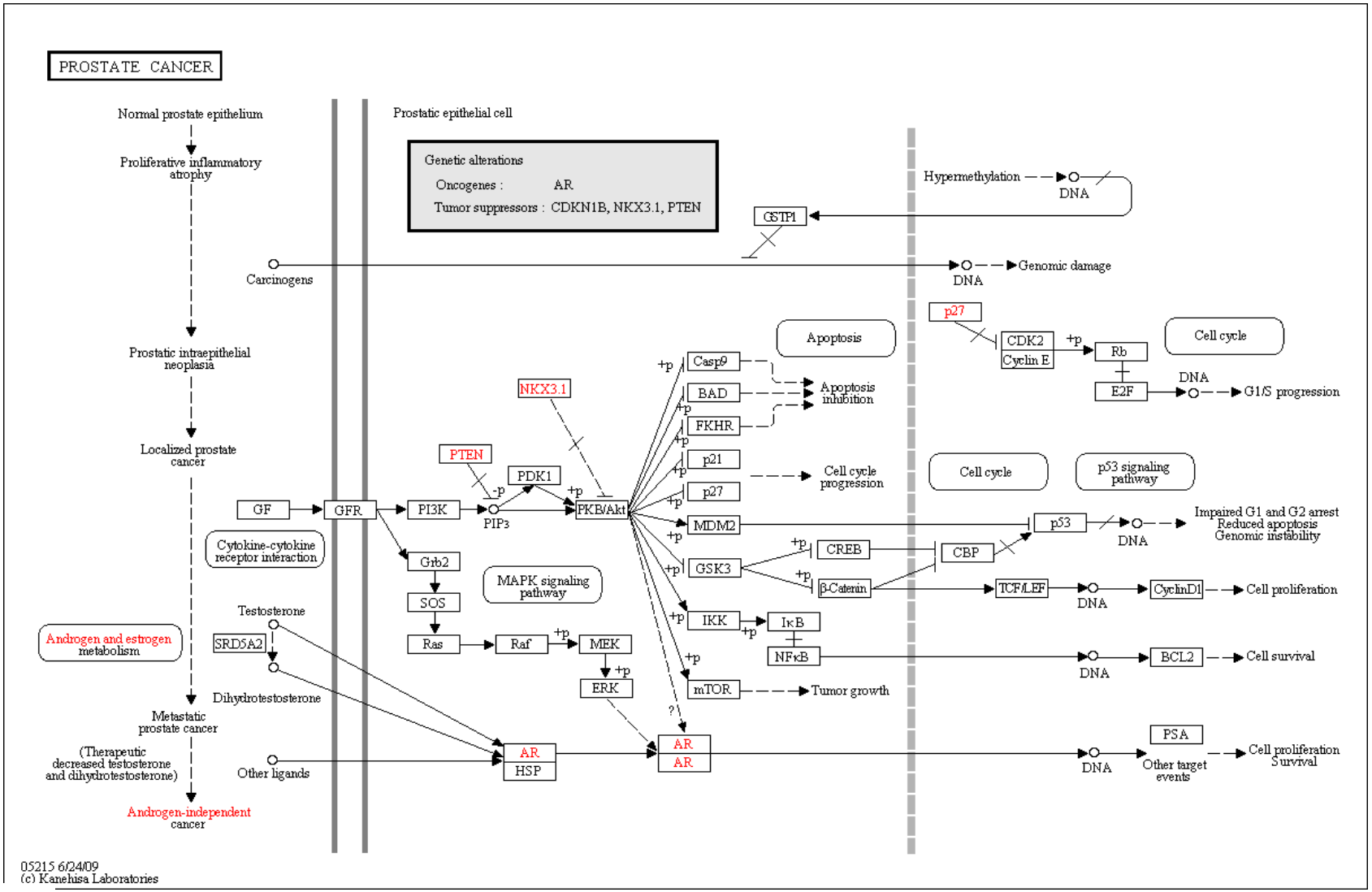
- Update j-th column of L at a time for j=(p-1) to 1.
- Randomly choose one of the acceptable move (Add zero, remove zero, or zero unchanged).
- Update the locations of zero and propose new values for nonzero locations based on a proposal distribution q.
- Accept the move with probability $\min\left\{1, \frac{\pi((J^*, W^*)|D) q_0(J^*, J)}{\pi((J, W)|D) q_0(J, J^*)}\right\}$

Application : Androgen pathway interaction

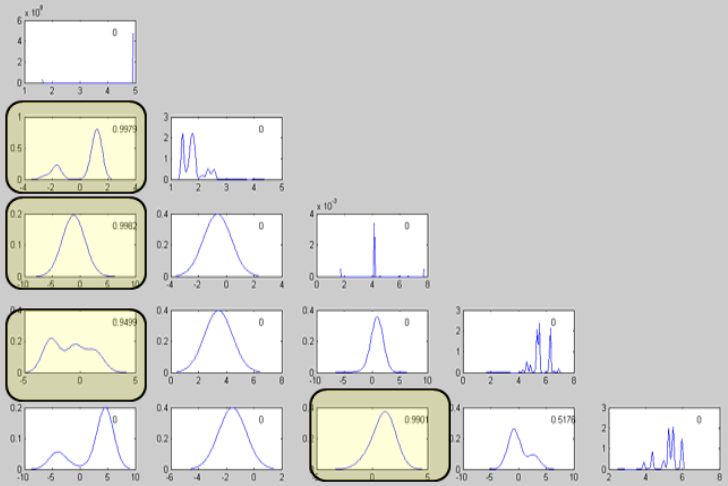
Peter J.M. Hendriksen et al , Cancer Res 2006

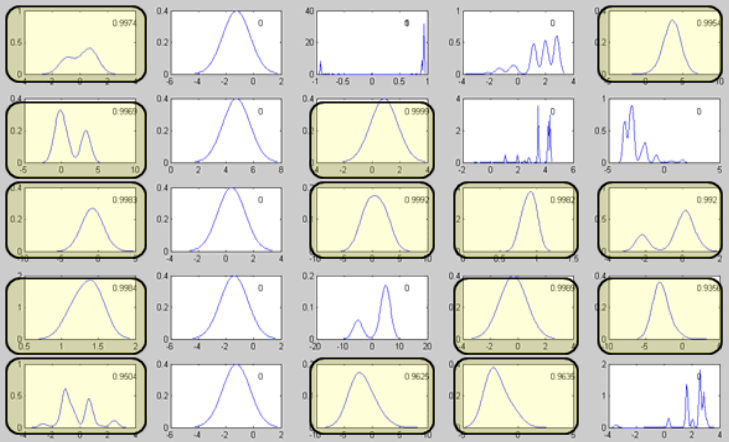
- **Gene expression omnibus :** GSE4084
- 12 samples on 10 genes.
- Genes are observed to be over expressed in cancer cells according to A.P. Singh et al. Cancer Letters 259 (2008)
- Marginal analysis for each gene is reported.
- Combined analysis of all genes will require ~100 covariance parameters.

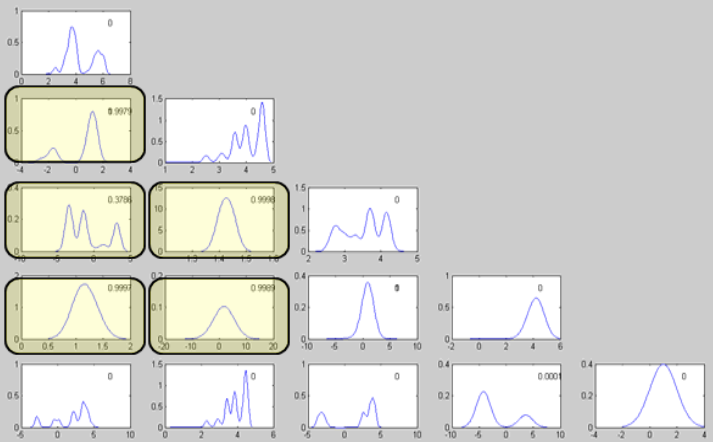
Kegg Pathway : Androgen



10,000 simulated samples of L :

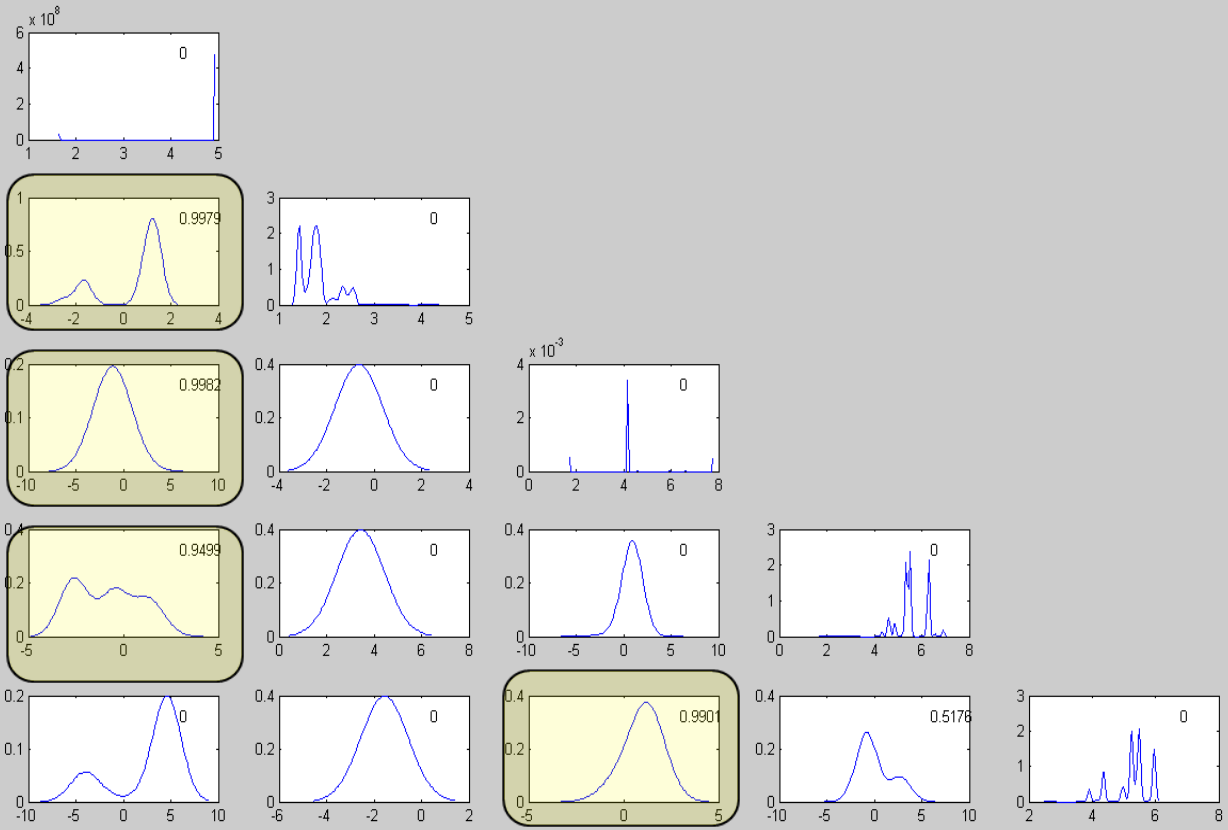




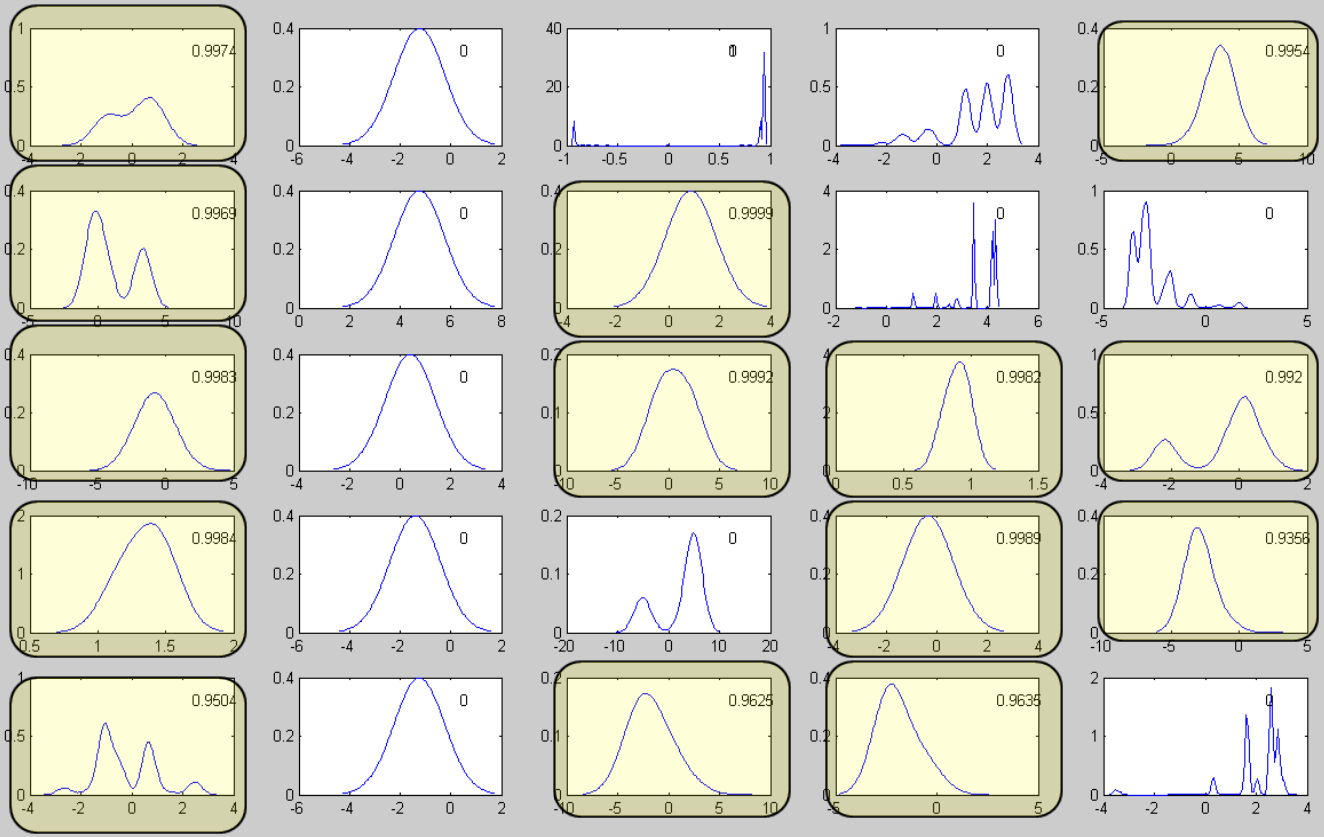


- Fitted density of 10000 nonzero values for each location.
- Actual distribution is a mixture of this distribution with a point mass at zero.
- Highlighted locations indicate large point mass on zero.

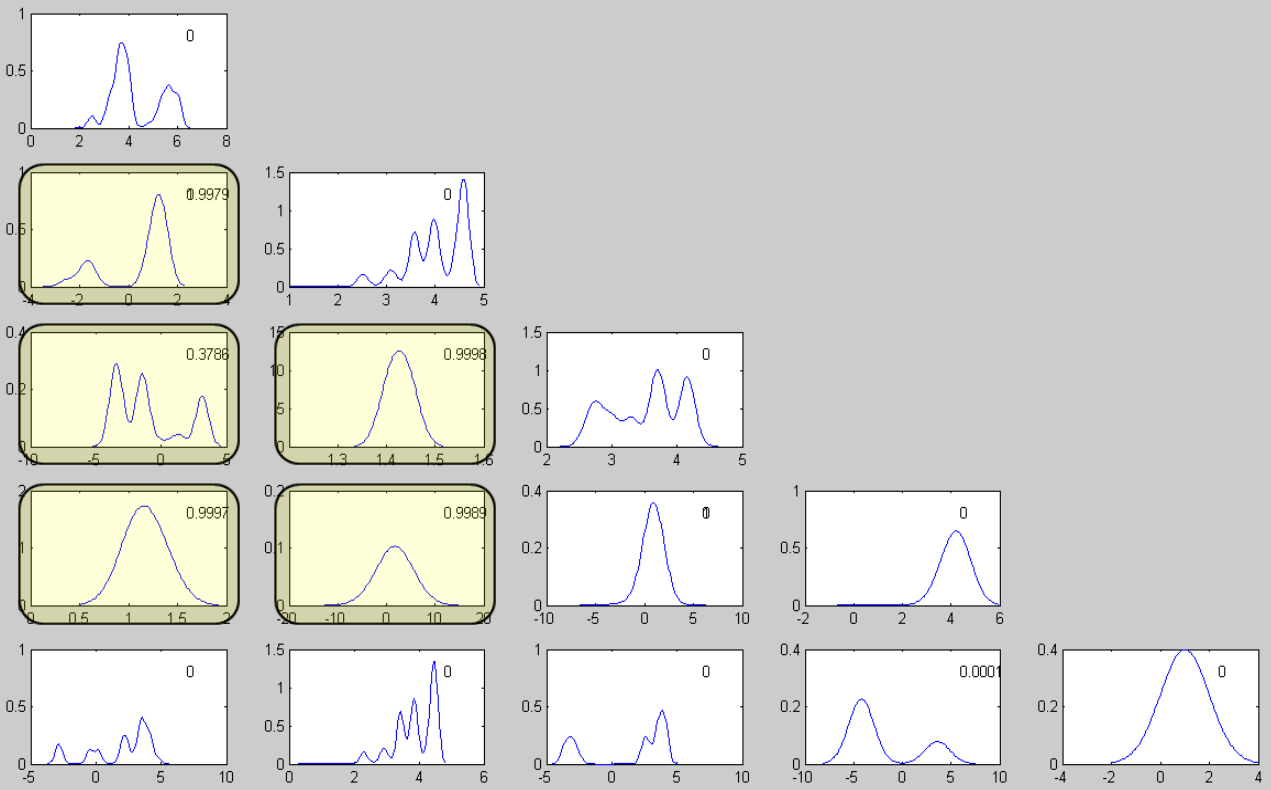
10,000 simulated samples of L :



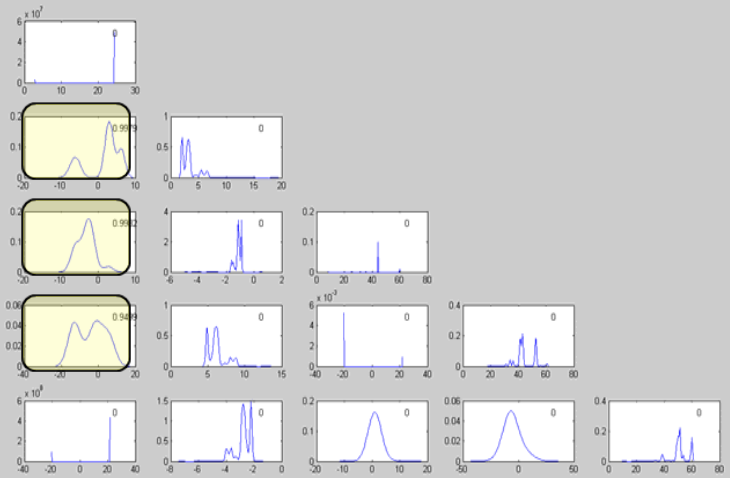
10,000 simulated samples of L :

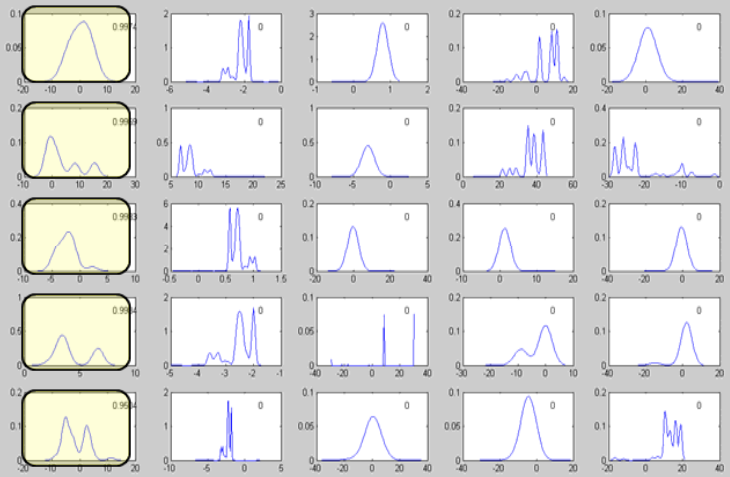


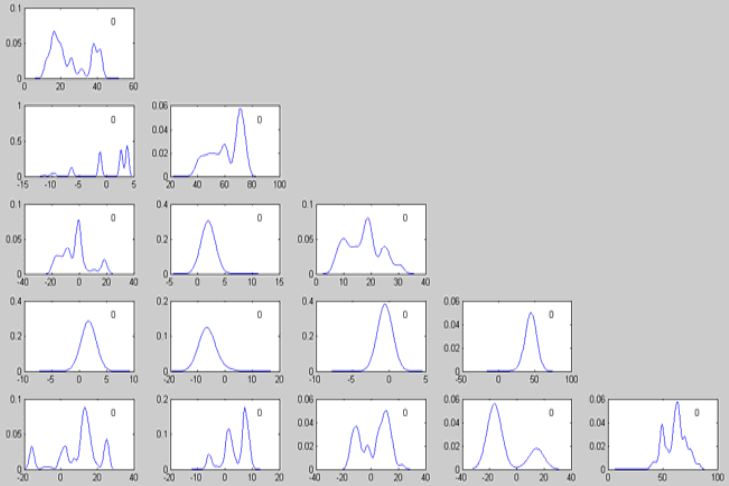
10,000 simulated samples of L :



10,000 simulated samples of W :

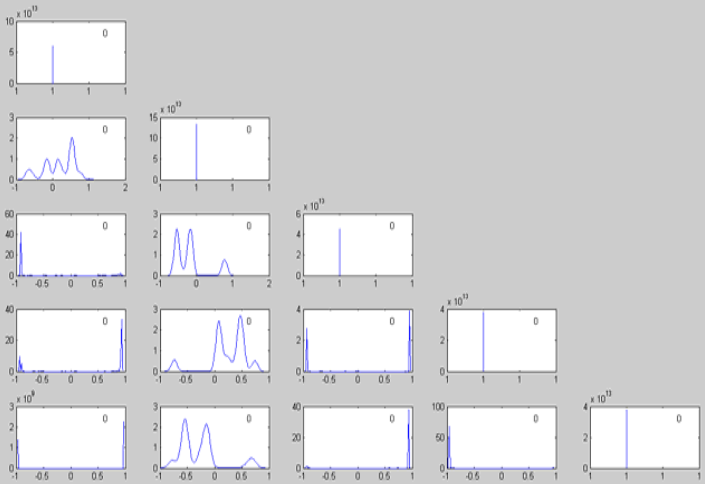




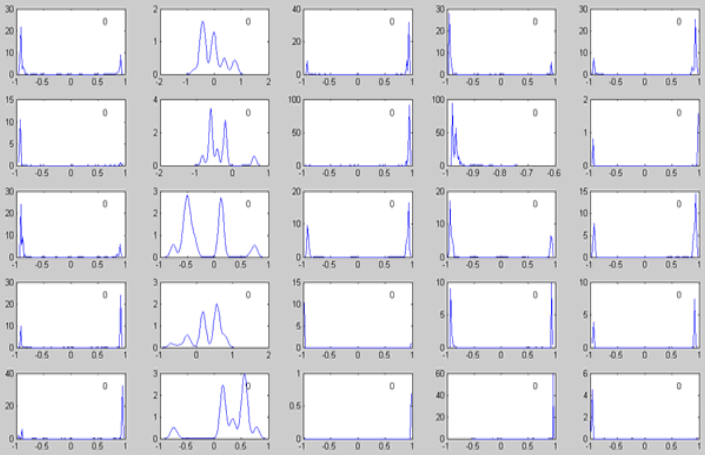


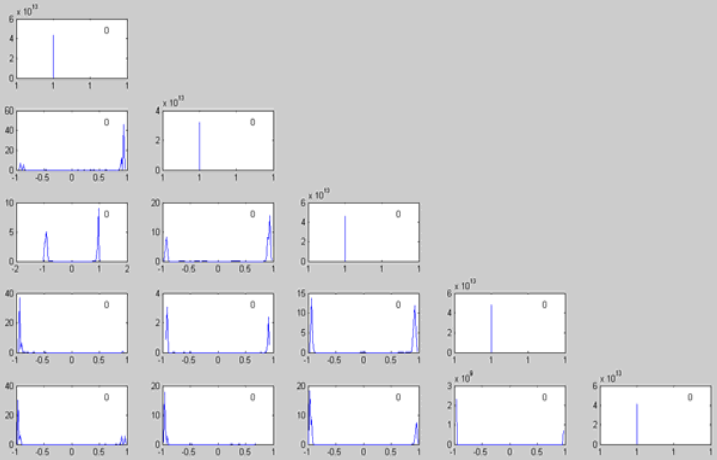
- Sparsity of L is transferred to W only
- for the first column.
- Other columns may not be sparse, but
- the generating probability is often lower dimensional.

10,000 simulated samples of R:



- Diagonals are exactly 1.
- Sparsity is not similar, but the underlying probability is lower dimensional.





Simulation with a known inverse correlation :

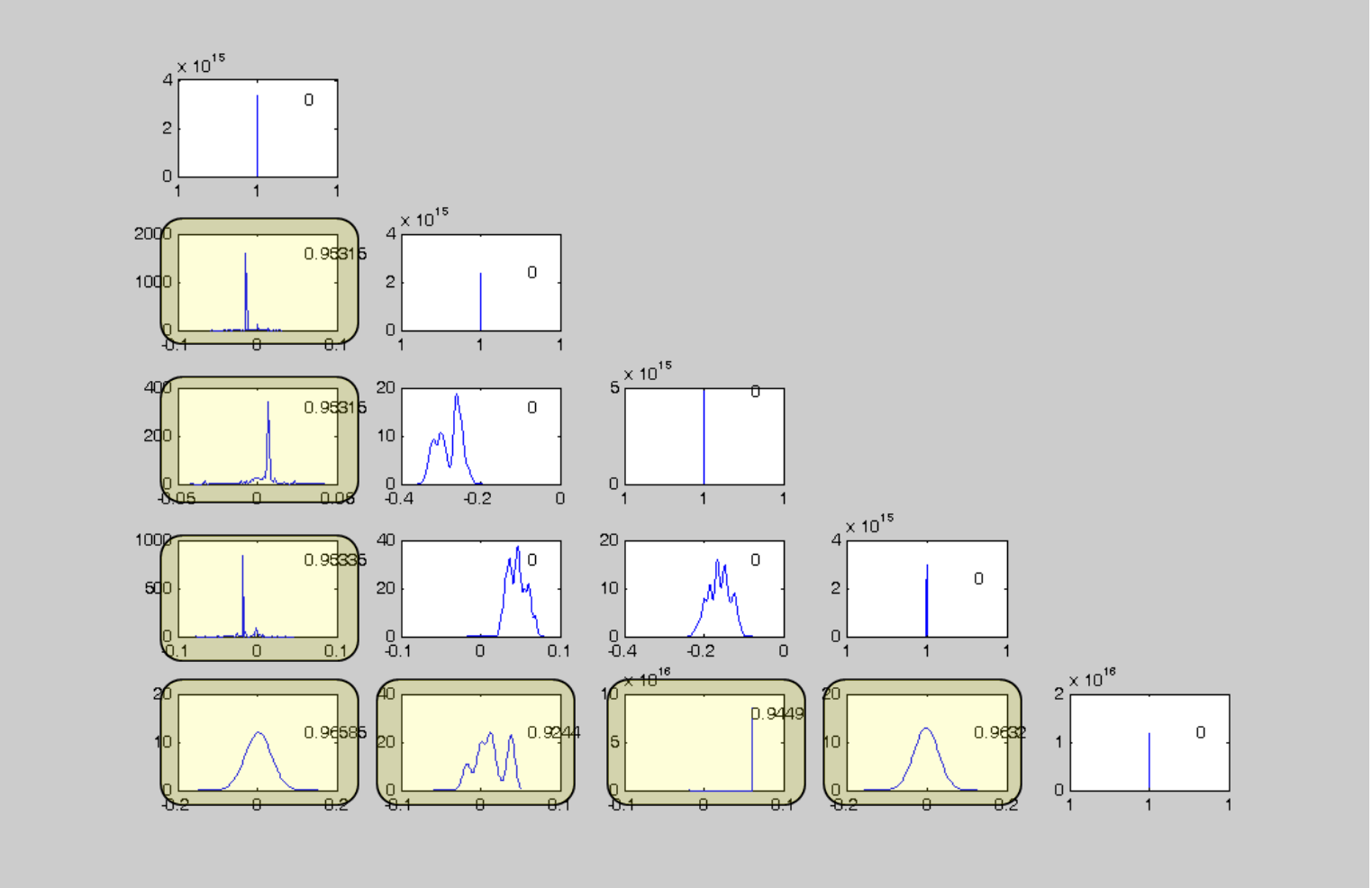
$$X \sim N(0, R) \; where,$$

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -0.28 & 0 & 0 \\ 0 & -0.28 & 1 & -0.2 & 0 \\ 0 & 0.06 & -0.2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, W = R^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1.09 & 0.31 & 0 & 0 \\ 0 & 0.31 & 1.13 & 0.2 & 0 \\ 0 & 0 & 0.2 & 1.04 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

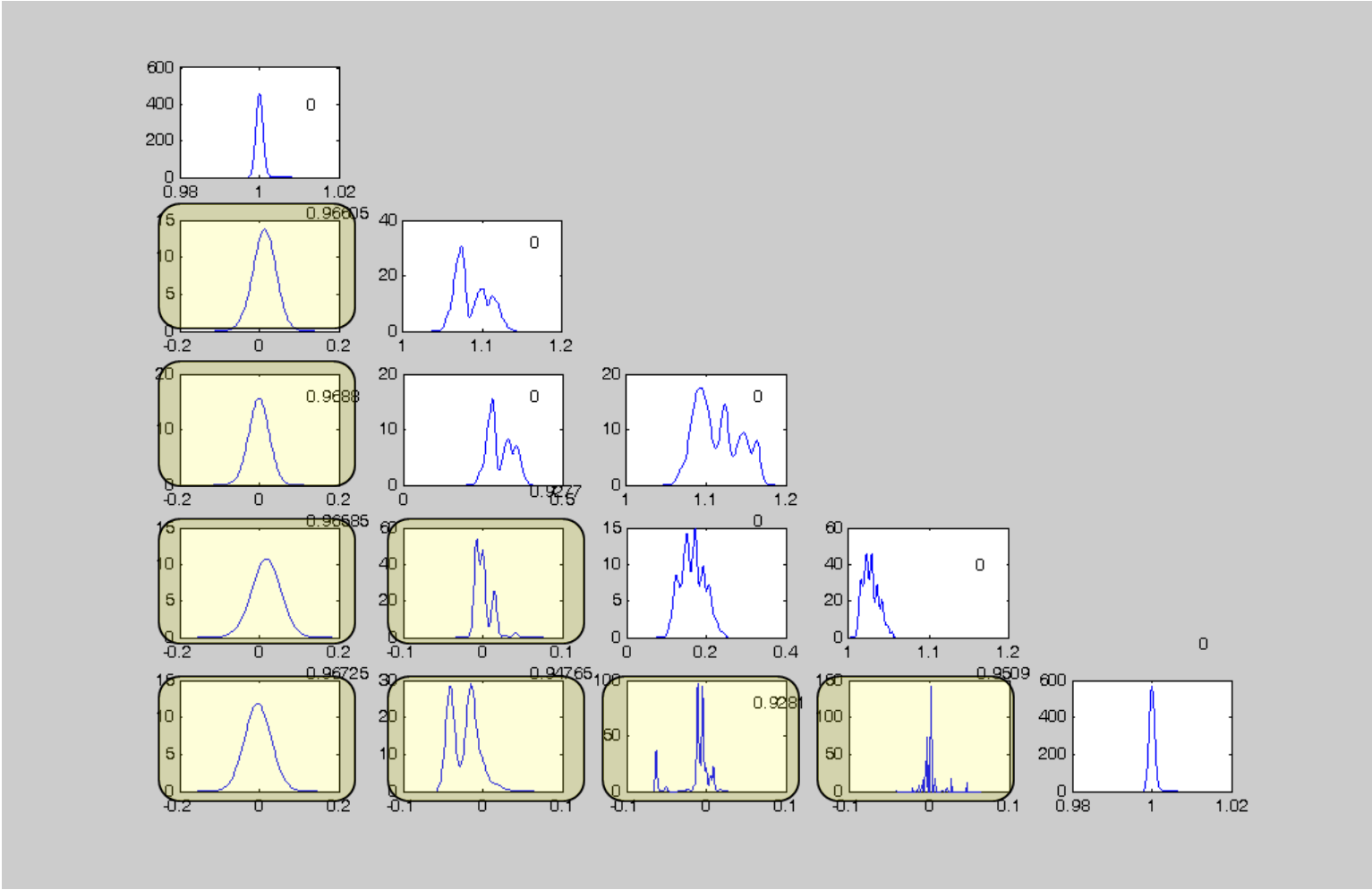
$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1.04 & 0 & 0 & 0 \\ 0 & 0.3 & 1.02 & 0 & 0 \\ 0 & 0 & 0.2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- Correlation and inverse correlation
- has a band matrix structure.
- Mostly sparse.
- Simulation has 200 observations,
- 20,000 MCMC samples.

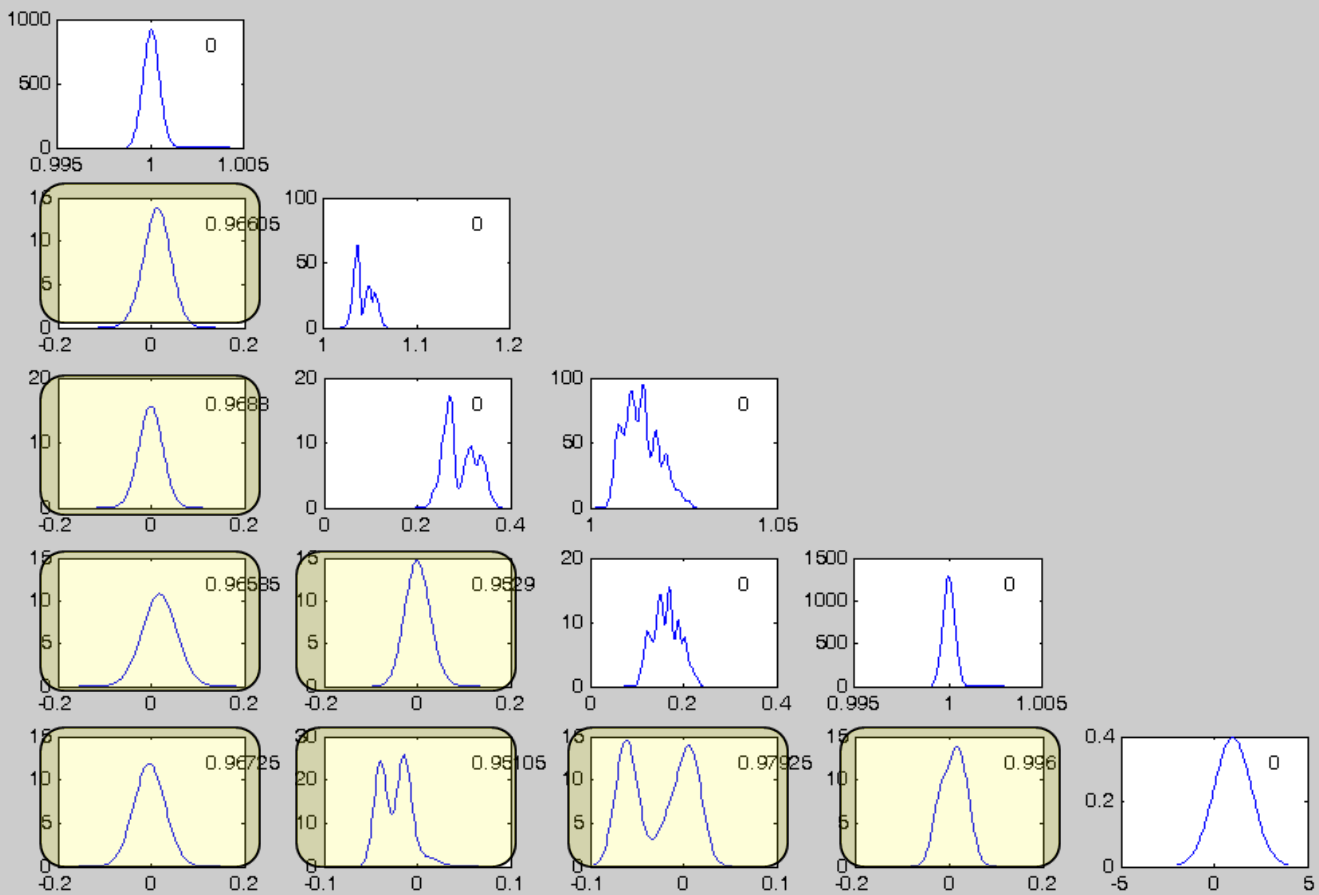
20,000 simulated samples of R



20,000 MCMC samples of W:



20,000 MCMC Samples of L :



Conclusion, Future directions :

- The algorithm makes good use of the sparse structure of the inverse correlation matrix.
- Easily generalized to much higher dimension, computation time would be higher depending on sparsity of the problem.
- Improves on updating scheme of updating one element at a time.
- We intend to apply this to graphical models without Gaussian structure by means of using copula to model the dependence.