

Covariance estimation via fiducial inference

Supplementary: Additional Simulations

W. Jenny Shi*

Computational Bioscience Program, University of Colorado e-mail: wjennyshi@gmail.com

Jan Hannig†

Department of Statistics & Operations Research, University of North Carolina e-mail: jan.hannig@unc.edu

Randy C.S. Lai

Department of Mathematics & Statistics, University of Maine e-mail: chushing.lai@maine.edu

Thomas C.M. Lee‡

Department of Statistics, University of California e-mail: tcmlee@ucdavis.edu

1. Clique example with Σ : $k = 8, p = 100, n = 1000$

Figure 1 shows the sparse structure estimation for a simulation with 10 chains started at random. The bottom right panel is based on the last 1000 iterations. The estimated cliques match true covariance exactly with fiducial probability one. Its corresponding confidence curve plots, along with the PDSCE estimator is shown in Figure 2. Left two panels show that the estimators peak very close to the truth. The right two panels indicate that the fiducial estimators are better than the sample covariance. Similar to previous examples, the PDSCE shows smaller distance to the true covariance matrix, but over estimates the log-determinant.

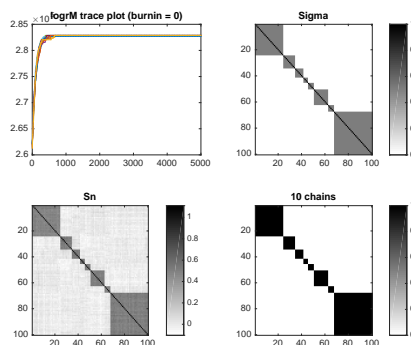
Repeat the simulation 200 times, each with one MCMC, we obtain the one-sided p-values. Its quantile-quantile plot against $\text{Unif}[0,1]$ is shown in Figure 3. The p-value curve (in green) weaves around the diagonal dotted guideline. It is well contained by the 95% coverage envelope, indicating good coverage.

*Shi's research was supported in part by the National Library of Medicine Institutional Training Grant T15LM009451

†Hannig's research was supported in part by the National Science Foundation (NSF) under Grant No. 1512945 and 1633074

‡Lee's research was supported in part by the NSF under Grant No. 1512945 and 1513484

Fig 1: Result for $k = 8, p = 100, n = 1000$. The trace plot (top left) shows that the chains converge quickly. Although $\frac{n}{p}$ is small, the sample covariance (bottom left) roughly captures the shape of true covariance (top right). The last panel (bottom right) shows that the fiducial estimate captures the true clique structure perfectly.



2. Σ with zero locations known: $p = 50, n = 50$

The confidence curves peak near the truth in the GFD panel, while slightly overestimates the covariance determinant. The PDSCE estimator overestimates the covariance determinant again. Our method and PDSCE produce estimators that are closer to the truth than the sample covariance. The last panel shows that the fiducial estimators have a relatively wide range of angles, yet the peaks are at least as good as the PDSCE estimator. For each Markov chain the burn in = 1000, and the window = 5000 (Figure 4).

3. General case: $p = 50, n = 50, \max C = 5$ (fixed zero locations unknown)

Working with the same true covariance matrix and observations as Example A2, we implement the algorithm for the general case where the sparse structure is unknown. The estimators fiducial are better than the sample covariance matrix. The PDSCE estimator appears to be slightly closer to the truth in FM-distance. Burn in = 50000, window = 10000 (See Figure 5).

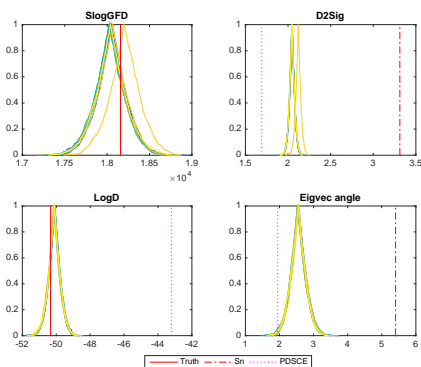


Fig 2: Confidence curve plots for estimated covariance matrix. $k = 8, p = 100, n = 1000$. Comparing to the sample covariance, the estimators are closer to Σ . The PDSCE estimator shows even smaller FM-distance to Σ , it, however, greatly overestimates $\det \Sigma$.

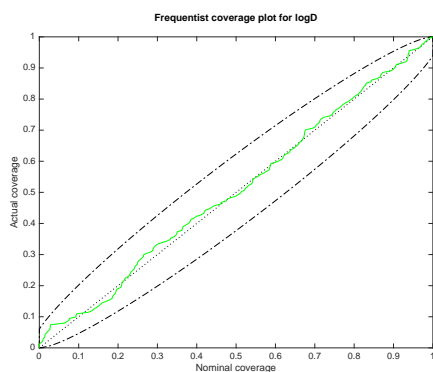


Fig 3: 95% coverage log-determinant plots for 200 repeated simulations. $k = 8, p = 100, n = 1000$. The p-values (in green) roughly follow a uniform $[0,1]$ distribution. It shows well calibrated coverage.

Fig 4: Overall, the fiducial estimators are slightly better than the PDSCE estimator in this example. They are both closer to the truth than the sample covariance matrix.

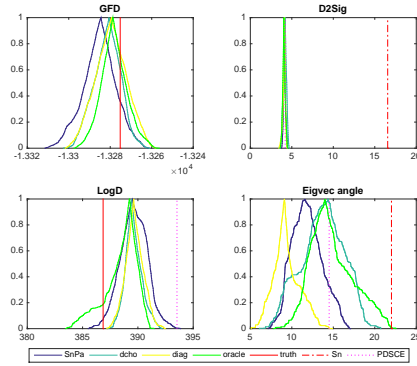


Fig 5: The fiducial estimators are better than the sample covariance matrix in the D2Sig and the Eigvec angle panels. The PDSCE estimator seems to be slightly better in those two panels, but overestimates the covariance determinant.

