Efficient Monte Carlo computation of Fisher information matrix using prior information

Sonjoy Das, UB
James C. Spall, APL/JHU
Roger Ghanem, USC

SIAM Conference on Data Mining
Anaheim, California, USA
April 26–28, 2012

Outline

1. **Introduction and Motivation**
   - General discussion of Fisher information matrix
   - Current resampling algorithm – No use of prior information
2. **Contribution/Results of the Proposed Work**
   - Improved resampling algorithm – using prior information
   - Theoretical basis
   - Numerical Illustrations

Outline

Outline

1. **Introduction and Motivation**
   - General discussion of Fisher information matrix
   - Current resampling algorithm – No use of prior information
2. **Contribution/Results of the Proposed Work**
   - Improved resampling algorithm – using prior information
   - Theoretical basis
   - Numerical Illustrations

Significance of Fisher Information Matrix

- Fundamental role of data analysis is to **extract information from data**
- Parameter estimation for models is central to process of extracting information
- The Fisher **information matrix** plays a central role in parameter estimation for measuring information:

Information matrix summarizes amount of information in data relative to parameters being estimated

Problem Setting

- Consider classical problem of estimating parameter vector, $\boldsymbol{\theta}$, from n data vectors, $\mathbf{Z}_n \equiv \{\mathcal{Z}_1, \cdots, \mathcal{Z}_n\}$
- Suppose have a probability density or mass function (PDF or PMF) associated with the data
- The parameter, $\boldsymbol{\theta}$, appears in the PDF or PMF and affect the nature of the distribution
  - Example: $\mathcal{Z}_i \sim N(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, for all i
- Let $\ell(\boldsymbol{\theta}|\mathbf{Z}_n)$ represents the likelihood function, *i.e.*, $\ell(\cdot)$ is the PDF or PMF viewed as a function of $\boldsymbol{\theta}$ conditioned on the data

Selected Applications

Information matrix is measure of performance for several applications. Five uses are:

1. **Confidence regions for parameter estimation**
   - Uses asymptotic normality and/or Cramér-Rao inequality
2. **Prediction bounds for mathematical models**
3. **Basis for "*D*-optimal" criterion for experimental design**
   - Information matrix serves as measure of how well θ can be estimated for a given set of inputs
4. **Basis for "noninformative prior" in Bayesian analysis**
   - Sometimes used for "objective" Bayesian inference
5. **Model selection**
   - Is model A "better" than model B?

Information Matrix

- Recall likelihood function, $\ell(\boldsymbol{\theta}|\mathbf{Z}_n)$ and the log-likelihood function by $L(\boldsymbol{\theta}|\mathbf{Z}_n) \equiv \ln \ell(\boldsymbol{\theta}|\mathbf{Z}_n)$
- Information matrix defined as

$$\boxed{\mathbf{F}_n(\boldsymbol{\theta}) \equiv E\left[\frac{\partial L}{\partial \boldsymbol{\theta}} \cdot \frac{\partial L}{\partial \boldsymbol{\theta}^T}\bigg| \boldsymbol{\theta}\right]}$$

  where expectation is w.r.t. the measure of $\mathbf{Z}_n$
- If Hessian matrix exists, equivalent form based on Hessian matrix:

$$\boxed{\mathbf{F}_n(\boldsymbol{\theta}) = -E\left[\frac{\partial^2 L}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^T}\bigg| \boldsymbol{\theta}\right]}$$

- $\mathbf{F}_n(\boldsymbol{\theta})$ is positive semidefinite of dimension $p \times p$, $p = \dim(\boldsymbol{\theta})$

Two Famous Results

Connection of $\mathbf{F}_n(\boldsymbol{\theta})$ and uncertainty in estimate, $\hat{\boldsymbol{\theta}}_n$, is rigorously specified via following results ($\boldsymbol{\theta}^*$ = true value of $\boldsymbol{\theta}$):

1. **Asymptotic normality:**

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\right) \xrightarrow{\text{dist}} N_p(\mathbf{0}, \bar{\mathbf{F}}^{-1})$$

where $\bar{\mathbf{F}} \equiv \lim_{n\to\infty} \mathbf{F}_n(\boldsymbol{\theta}^*)/n$

2. **Cramér-Rao inequality:**

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}_n) \geq \mathbf{F}_n(\boldsymbol{\theta}^*)^{-1}, \quad \text{for all } n \text{ (unbiased } \hat{\boldsymbol{\theta}}_n)$$

Above two results indicate: greater variability in $\hat{\boldsymbol{\theta}}_n \Longrightarrow$ "smaller" $\mathbf{F}_n(\boldsymbol{\theta})$ (and vice versa)

Outline

1. **Introduction and Motivation**
   - General discussion of Fisher information matrix
   - Current resampling algorithm – No use of prior information
2. **Contribution/Results of the Proposed Work**
   - Improved resampling algorithm – using prior information
   - Theoretical basis
   - Numerical Illustrations

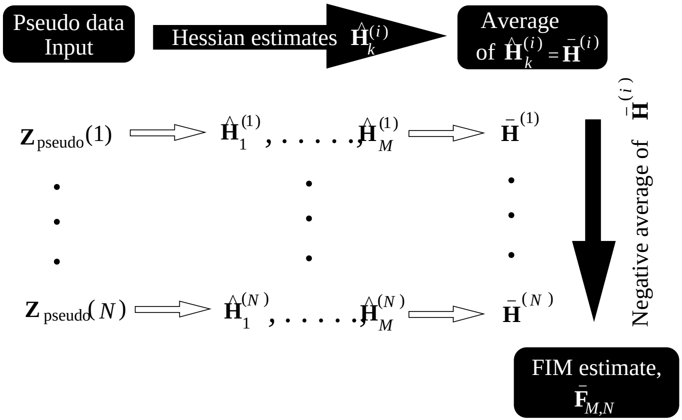
Monte Carlo Computation of Information Matrix

- Analytical formula for $\mathbf{F}_n(\theta)$ requires first or second derivative and expectation calculation
  - Often impossible or very difficult to compute in practical applications
  - Involves expected value of highly nonlinear (possibly unknown) functions of data
- Schematic next summarizes “easy” Monte Carlo-based method for determining $\mathbf{F}_n(\theta)$
  - Uses averages of very efficient (simultaneous perturbation) Hessian estimates
  - Hessian estimates evaluated at artificial (pseudo) data
  - Computational horsepower instead of analytical analysis

Monte Carlo Computation of Information Matrix

- Analytical formula for $\mathbf{F}_n(\theta)$ requires first or second derivative and expectation calculation
  - Often impossible or very difficult to compute in practical applications
  - Involves expected value of highly nonlinear (possibly unknown) functions of data
- Schematic next summarizes “easy” Monte Carlo-based method for determining $\mathbf{F}_n(\theta)$
  - Uses averages of very efficient (simultaneous perturbation) Hessian estimates
  - Hessian estimates evaluated at artificial (pseudo) data
  - Computational horsepower instead of analytical analysis

Schematic of Monte Carlo Method for Estimating Information Matrix



(Spall, 2005, *JCGS*)

Supplement: Simultaneous Perturbation (SP) Hessian and Gradient Estimate

$$\hat{\mathbf{H}}_k^{(i)} = \frac{1}{2}\left\{ \frac{\delta\mathbf{G}_k^{(i)}}{2\,c}\left[\Delta_{k1}^{-1},\cdots,\Delta_{kp}^{-1}\right] + \left(\frac{\delta\mathbf{G}_k^{(i)}}{2\,c}\left[\Delta_{k1}^{-1},\cdots,\Delta_{kp}^{-1}\right]\right)^T \right\}$$

where

$$\delta\mathbf{G}_k^{(i)} \equiv \mathbf{G}(\boldsymbol{\theta}+c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(i)) - \mathbf{G}(\boldsymbol{\theta}-c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(i))$$

$$\mathbf{G}(\boldsymbol{\theta}\pm c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(i)) = \frac{\partial L(\boldsymbol{\theta}\pm c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(i))}{\partial\boldsymbol{\theta}}$$

OR

$$= (1/\tilde{c})\left[L(\boldsymbol{\theta}+\tilde{c}\tilde{\boldsymbol{\Delta}}_k \pm c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(i)) - L(\boldsymbol{\theta}\pm c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(i))\right]\begin{bmatrix}\tilde{\Delta}_{k1}^{-1}\\ \vdots \\ \tilde{\Delta}_{kp}^{-1}\end{bmatrix}$$

- $\boldsymbol{\Delta}_k = [\Delta_{k1},\cdots,\Delta_{kp}]^T$ and $\Delta_{k1},\cdots,\Delta_{kp}$, mean-zero and statistically independent r.v.s with finite inverse moments
- $\tilde{\boldsymbol{\Delta}}_k$ has same statistical properties as $\boldsymbol{\Delta}_k$
- $\tilde{c} > c > 0$ are “small’ numbers

Supplement: Simultaneous Perturbation (SP) Hessian and Gradient Estimate

$$\hat{\mathbf{H}}_k^{(i)} = \frac{1}{2}\left\{ \frac{\delta \mathbf{G}_k^{(i)}}{2\,c}\left[\Delta_{k1}^{-1},\cdots,\Delta_{kp}^{-1}\right] + \left(\frac{\delta \mathbf{G}_k^{(i)}}{2\,c}\left[\Delta_{k1}^{-1},\cdots,\Delta_{kp}^{-1}\right]\right)^T\right\}$$

where

$$\delta\mathbf{G}_k^{(i)} \equiv \mathbf{G}(\boldsymbol{\theta} + c\,\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(i)) - \mathbf{G}(\boldsymbol{\theta} - c\,\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(i))$$

$$\mathbf{G}(\boldsymbol{\theta} \pm c\,\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(i)) = \frac{\partial L(\boldsymbol{\theta} \pm c\,\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(i))}{\partial \boldsymbol{\theta}}$$

OR

$$= (1/\tilde{c})\left[L(\boldsymbol{\theta} + \tilde{c}\,\tilde{\boldsymbol{\Delta}}_k \pm c\,\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(i)) - L(\boldsymbol{\theta} \pm c\,\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(i))\right]\begin{bmatrix}\tilde{\Delta}_{k1}^{-1}\\ \vdots \\ \tilde{\Delta}_{kp}^{-1}\end{bmatrix}$$

- $\boldsymbol{\Delta}_k = [\Delta_{k1},\cdots,\Delta_{kp}]^T$ and $\Delta_{k1},\cdots,\Delta_{kp}$, mean-zero and statistically independent r.v.s with finite inverse moments
- $\tilde{\boldsymbol{\Delta}}_k$ has same statistical properties as $\boldsymbol{\Delta}_k$
- $\tilde{c} > c > 0$ are “small’ numbers

Supplement: Optimal Implementation

Several implementation questions/answers:

Q. How to compute (cheap) Hessian estimates?

A. Use simultaneous perturbation (SP) based method (Spall, 2000, *IEEE Trans. Auto. Control*)

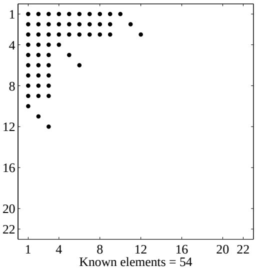
Q. How to allocate per-realization (M) and across-realization (N) averaging?

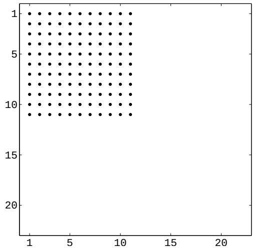
A. $M = 1$ is the optimal solution for a fixed total number of Hessian estimates. However, $M > 1$ is useful when accounting for cost of generating pseudo data

Q. Can correlation be introduced to improve overall accuracy of $\bar{\mathbf{F}}_{M,N}(\theta)$?

A. Yes, ***antithetic random numbers*** can reduce variance of elements in $\bar{\mathbf{F}}_{M,N}(\theta)$ (Spall, 2005, *JCGS*)

Fisher information matrix with analytically known elements





(Das, Ghanem, Spall, 2008, *SISC*)

(Navigation application at APL)

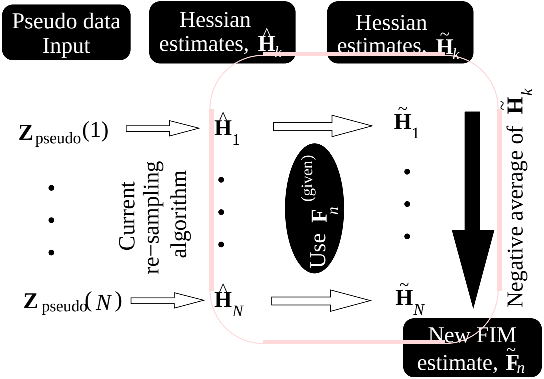
The previous resampling approach (Spall, 2005, *JCGS*) yields the "full" Fisher information matrix **without considering the available prior information**

Outline

1. Introduction and Motivation
   - General discussion of Fisher information matrix
   - Current resampling algorithm – No use of prior information

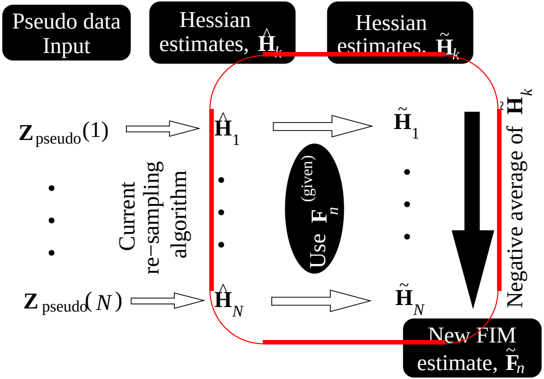
2. Contribution/Results of the Proposed Work
   - Improved resampling algorithm – using prior information
   - Theoretical basis
   - Numerical Illustrations

Schematic of the Proposed Resampling Algorithm



For $M = 1$; however, can be readily extended to the case when $M > 1$

Schematic of the Proposed Resampling Algorithm



For $M = 1$; however, can be readily extended to the case when $M > 1$

Outline

1. Introduction and Motivation
   - General discussion of Fisher information matrix
   - Current resampling algorithm – No use of prior information

2. Contribution/Results of the Proposed Work
   - Improved resampling algorithm – using prior information
   - Theoretical basis
   - Numerical Illustrations

Supplement: Improvement in terms of Variance Reduction

Case 1: Log-likelihood based

$$\begin{aligned}&\text{var}[\hat{H}_{ii}^{(L)}|\boldsymbol{\theta}] - \text{var}[\tilde{H}_{ii}^{(L)}|\boldsymbol{\theta}] \\ \approx &\sum_{l}\sum_{m\in\mathbb{I}_l} a_{lm}(i,i)\,(F_{lm}(\boldsymbol{\theta}))^2 \\ > &\ 0, \quad i \in \mathbb{I}_i^c\end{aligned}$$

where
$a_{lm}(i,j) = E[\Delta_m^2/\Delta_j^2]E[\tilde{\Delta}_l^2/\tilde{\Delta}_i^2] > 0$

Case 2: Gradient based

$$\begin{aligned}&\text{var}[\hat{H}_{ii}^{(\mathbf{g})}|\boldsymbol{\theta}] - \text{var}[\tilde{H}_{ii}^{(\mathbf{g})}|\boldsymbol{\theta}] \\ \approx &\sum_{l\in\mathbb{I}_i} b_l(i)\,(F_{il}(\boldsymbol{\theta}))^2 \\ > &\ 0, \quad i \in \mathbb{I}_i^c\end{aligned}$$

where $b_l(j) = E[\Delta_l^2/\Delta_j^2] > 0$

- For (i,j)-th element an additional covariance terms needed to be considered and it can still be shown that $\text{var}[\tilde{H}_{ij}|\boldsymbol{\theta}] < \text{var}[\hat{H}_{ij}|\boldsymbol{\theta}]$ (see the paper)
- $\Longrightarrow$ $\text{var}[\tilde{F}_{ij}|\boldsymbol{\theta}] < \text{var}[\hat{F}_{ij}|\boldsymbol{\theta}]$
- Also, $E[\tilde{H}_{ij}|\boldsymbol{\theta}] = E[\hat{H}_{ij}|\boldsymbol{\theta}]$, for all $j \in \mathbb{I}_i^c$, for both cases

Basic Ideas for Proofs/Implementation

- Per current resampling algorithm, the (i,j)-th element is given by,

$$\hat{H}_{ij} \approx \sum_{\text{unknown+known}} \omega_{lm} H_{lm}, \quad \text{(weighted sum of **unknown** elements of } \mathbf{H}_k)$$

$$\Rightarrow \mathrm{var}(\hat{H}_{ij}) \approx E\Big[\Big(\sum_{\text{unknown+known}} \omega_{lm} H_{lm}\Big)^2\Big] - (F_{ij}(\boldsymbol{\theta}))^2, \quad \text{since } E[\hat{H}_{ij}|\boldsymbol{\theta}] \approx -F_{ij}(\boldsymbol{\theta})$$

- Per modified resampling algorithm,

$$\hat{H}_{ij} \approx \sum \omega_{lm} H_{lm} + \text{parts corresponding to unknown elements of } \mathbf{F}_n(\boldsymbol{\theta})$$

$$H_{lm} = -F_{lm} + e_{lm}, \quad e_{lm} \equiv \text{mean-zero error terms}$$

  - The new estimate, therefore, is given by,

$$\tilde{H}_{ij} \equiv \Big[\hat{H}_{ij} - \sum_{\text{known}} \omega_{lm}(-F_{lm})\Big] \approx \sum_{\text{unknown}} \omega_{lm} H_{lm} + \sum_{\text{known}} \omega_{lm} e_{lm}$$

$$\Rightarrow \mathrm{var}(\tilde{H}_{ij}) \approx E\Big[\Big(\sum_{\text{unknown}} \omega_{lm} H_{lm} + \sum_{\text{known}} \omega_{lm} e_{lm}\Big)^2\Big] - (F_{ij}(\boldsymbol{\theta}))^2$$

  - Since $E[e_{lm}^2] \equiv \mathrm{var}(H_{lm}) < E[H_{lm}^2] \Rightarrow \mathrm{var}(\tilde{H}_{ij}) < \mathrm{var}(\hat{H}_{ij})$

    - $\Longrightarrow$ $\mathrm{var}[\tilde{F}_{ij}|\boldsymbol{\theta}] < \mathrm{var}[\hat{F}_{ij}|\boldsymbol{\theta}]$

Basic Ideas for Proofs/Implementation

- Per current resampling algorithm, the (i,j)-th element is given by,

$$\hat{H}_{ij} \approx \sum_{\text{unknown+known}} \omega_{lm} H_{lm}, \quad \text{(weighted sum of } \textbf{unknown} \text{ elements of } \mathbf{H}_k\text{)}$$

$$\Rightarrow \mathrm{var}(\hat{H}_{ij}) \approx E\Big[\Big(\sum_{\text{unknown+known}} \omega_{lm} H_{lm}\Big)^2\Big] - (F_{ij}(\boldsymbol{\theta}))^2, \quad \text{since } E[\hat{H}_{ij}|\boldsymbol{\theta}] \approx -F_{ij}(\boldsymbol{\theta})$$

- Per modified resampling algorithm,

$$\hat{H}_{ij} \approx \sum \omega_{lm} H_{lm} + \text{parts corresponding to unknown elements of } \mathbf{F}_n(\boldsymbol{\theta})$$

$$H_{lm} = -F_{lm} + e_{lm}, \quad e_{lm} \equiv \textbf{mean-zero error terms}$$

  - The new estimate, therefore, is given by,

$$\tilde{H}_{ij} \equiv \Big[\hat{H}_{ij} - \sum_{\text{known}} \omega_{lm}(-F_{lm})\Big] \approx \sum_{\text{unknown}} \omega_{lm} H_{lm} + \sum_{\text{known}} \omega_{lm} e_{lm}$$

$$\Rightarrow \mathrm{var}(\tilde{H}_{ij}) \approx E\Big[\Big(\sum_{\text{unknown}} \omega_{lm} H_{lm} + \sum_{\text{known}} \omega_{lm} e_{lm}\Big)^2\Big] - (F_{ij}(\boldsymbol{\theta}))^2$$

  - Since $E[e_{lm}^2] \equiv \mathrm{var}(H_{lm}) < E[H_{lm}^2] \Rightarrow \mathrm{var}(\tilde{H}_{ij}) < \mathrm{var}(\hat{H}_{ij})$
    - $\Longrightarrow$ $\mathrm{var}[\tilde{F}_{ij}|\boldsymbol{\theta}] < \mathrm{var}[\hat{F}_{ij}|\boldsymbol{\theta}]$

Basic Ideas for Proofs/Implementation

- Per current resampling algorithm, the (i, j)-th element is given by,

$$\hat{H}_{ij} \approx \sum_{\text{unknown+known}} \omega_{lm} H_{lm}, \quad \text{(weighted sum of } \textbf{unknown} \text{ elements of } \mathbf{H}_k\text{)}$$

$$\Rightarrow \text{var}(\hat{H}_{ij}) \approx E\Big[\Big(\sum_{\text{unknown+known}} \omega_{lm} H_{lm}\Big)^2\Big] - (F_{ij}(\boldsymbol{\theta}))^2, \quad \text{since } E[\hat{H}_{ij}|\boldsymbol{\theta}] \approx -F_{ij}(\boldsymbol{\theta})$$

- Per modified resampling algorithm,

$$\hat{H}_{ij} \approx \sum \omega_{lm} \cancel{H}_{lm} + \text{parts corresponding to unknown elements of } \mathbf{F}_n(\boldsymbol{\theta})$$

$$H_{lm} = -F_{lm} + e_{lm}, \quad e_{lm} \equiv \textbf{mean-zero error terms}$$

  - The new estimate, therefore, is given by,

$$\tilde{H}_{ij} \equiv \Big[\hat{H}_{ij} - \sum_{\text{known}} \omega_{lm}(-F_{lm})\Big] \approx \sum_{\text{unknown}} \omega_{lm} H_{lm} + \sum_{\text{known}} \omega_{lm} e_{lm}$$

$$\Rightarrow \text{var}(\tilde{H}_{ij}) \approx E\Big[\Big(\sum_{\text{unknown}} \omega_{lm} H_{lm} + \sum_{\text{known}} \omega_{lm} e_{lm}\Big)^2\Big] - (F_{ij}(\boldsymbol{\theta}))^2$$

  - Since $E[e_{lm}^2] \equiv \text{var}(H_{lm}) < E[H_{lm}^2] \Rightarrow \text{var}(\tilde{H}_{ij}) < \text{var}(\hat{H}_{ij})$
    - $\Longrightarrow$ $\boxed{\text{var}[\tilde{F}_{ij}|\boldsymbol{\theta}] < \text{var}[\hat{F}_{ij}|\boldsymbol{\theta}]}$

Outline

1. Introduction and Motivation
   - General discussion of Fisher information matrix
   - Current resampling algorithm – No use of prior information

2. Contribution/Results of the Proposed Work
   - Improved resampling algorithm – using prior information
   - Theoretical basis
   - Numerical Illustrations

Problem Description

- Consider independently distributed scalar-valued random data $\mathbf{z}_i \sim N(\mu, \sigma^2 + c_i\alpha)$, $\forall i$, $n = 30$
  - A problem with known information matrix
  - Useful for comparing simulation results with known analytical results
  - $\boldsymbol{\theta} = [\mu, \sigma^2, \alpha]^T$ and assume $\mu = 0$, $\sigma^2 = 1$ and $\alpha = 1$ for the purpose of illustration
  - $0 < c_i < 1$ assumed known (non-identical across i)
- $p = \dim(\boldsymbol{\theta}) = 3$

  $\Longrightarrow 3(3+1)/2 = 6$ unique elements in $\mathbf{F}_n(\boldsymbol{\theta})$
  - Assume that only the **upper-left 2×2 block** of $\mathbf{F}_n(\boldsymbol{\theta})$ **known *a priori***
- The analytical Fisher information matrix in practical applications is **not** known (unlike this example) or partially known

Results

Analytical FIM

$$\mathbf{F}_n(\boldsymbol{\theta}) = \begin{bmatrix} F_{11} & 0 & 0 \\ 0 & F_{22} & F_{23} \\ 0 & F_{23} & F_{33} \end{bmatrix}$$

For illustration,

$$\mathbf{F}_n^{\text{given}}(\boldsymbol{\theta}) = \begin{bmatrix} F_{11} & 0 & ? \\ 0 & F_{22} & ? \\ ? & ? & ? \end{bmatrix},$$

	Error in FIM estimates		MSE
	relMSE($\hat{\mathbf{F}}_n$) [MSE($\hat{\mathbf{F}}_n$)]	relMSE($\tilde{\mathbf{F}}_n$) [MSE($\tilde{\mathbf{F}}_n$)]	(variance) reduction
L-based	0.00135 % [0.0071]	0.00011 % [0.0006]	91.5 % (93.4 %)
$\mathbf{g}$-based	0.0533 % [0.0020]	0.0198 % [0.0004]	81.0 % (93.5 %)

MSE and MSE reduction of estimates for $\mathbf{F}_n(\boldsymbol{\theta})$,
$N = 5 \times 10^5$, $c = 0.0001$, $\tilde{c} = 0.00011$,
$\Delta_{ki}, \tilde{\Delta}_{ki} \sim \text{Bernoulli}\pm 1$

Concluding Remarks

- Fisher information matrix is a central quantity in data analysis and parameter estimation
  - Direct computation of information matrix in general nonlinear problems usually impossible
  - Monte Carlo approach usually preferred
- A **modification** of the previous Monte Carlo based resampling algorithm is proposed to **enhance the statistical characteristics** of the estimator of $\mathbf{F}_n(\theta)$
  - Particularly **useful** in those cases where **some elements** of $\mathbf{F}_n(\theta)$ are **analytically known from prior information**
  - Numerical illustrations show considerable improvement of the new estimator (in the sense of **mean-squared error reduction as well as variance reduction**) over the previous estimator

Concluding Remarks

- Fisher information matrix is a central quantity in data analysis and parameter estimation
  - Direct computation of information matrix in general nonlinear problems usually impossible
  - Monte Carlo approach usually preferred
- A **modification** of the previous Monte Carlo based resampling algorithm is proposed to **enhance the statistical characteristics** of the estimator of $\mathbf{F}_n(\theta)$
  - Particularly **useful** in those cases where **some elements** of $\mathbf{F}_n(\theta)$ are **analytically known from prior information**
  - Numerical illustrations show considerable improvement of the new estimator (in the sense of **mean-squared error reduction as well as variance reduction**) over the previous estimator

For Further Reading

- S. Das, “Efficient calculation of Fisher information matrix: Monte Carlo approach using prior information,” Master’s thesis, Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, Maryland, USA, May 2007, `http://dspace.library.jhu.edu/handle/1774.2/32459`.
- J. C. Spall, “Monte carlo computation of the Fisher information matrix in nonstandard settings,” *J. Comput. Graph. Statist.*, vol. 14, no. 4, pp. 889–909, 2005.
- S. Das, J. C. Spall, and R. Ghanem, “Efficient Monte Carlo computation of Fisher information matrix using prior information,” *Computational Statistics and Data Analysis*, Volume 54, No. 2, pp. 272–289, 2010.