EFFICIENT CALCULATION OF FISHER
INFORMATION MATRIX: MONTE CARLO APPROACH
USING PRIOR INFORMATION

by

Sonjoy Das

A thesis submitted to The Johns Hopkins University in conformity with the
requirements for the degree of Master of Science in Engineering.

Baltimore, Maryland
May, 2007

© Sonjoy Das 2007
All rights reserved
Abstract

The Fisher information matrix (FIM) is a critical quantity in several aspects of mathematical modeling, including input selection and confidence region calculation. Analytical determination of the FIM in a general setting, specially in nonlinear models, may be difficult or almost impossible due to intractable modeling requirements and/or intractable high-dimensional integration.

To circumvent these difficulties, a Monte Carlo (MC) simulation-based technique, resampling algorithm, based on the values of log-likelihood function or its exact stochastic gradient computed by using a set of pseudo data vectors, is usually recommended. The current work proposes an extension of the resampling algorithm in order to enhance the statistical qualities of the estimator of the FIM. This modified resampling algorithm is useful in those cases where the FIM has a structure with some elements being analytically known from prior information and the others being unknown. The estimator of the FIM, obtained by using the proposed algorithm, simultaneously preserves the analytically known elements and reduces variances of the estimators of the unknown elements by capitalizing on information contained in the
known elements. Numerical illustrations considered in this work show considerable improvement of the new FIM estimator (in the sense of mean-squared error reduction as well as variance reduction) based on the modified MC based resampling algorithm over the FIM estimator based on the current MC based resampling algorithm.

Advisor: Prof James C. Spall. The Johns Hopkins University.
Acknowledgements

I would like to express my sincere gratitude and appreciation to my advisor, Professor James C. Spall, for his insightful suggestions, technical guidance, patience, support and encouragement throughout the course of this work. I also like to formally express my sincere thanks and appreciation to the advisor of my doctoral program, Professor Roger Ghanem, at the Department of Civil and Environmental Engineering, University of Southern California, for his constant support and allowing me to continue this work within the already busy schedule of my doctoral work, without which this work would definitely have not been completed. In addition, I thank Professor Daniel Q. Naiman for his support to my endeavor for the successful completion of this work.

I am thankful to Ms. Kristin Bechtel, Academic Program Coordinator of the department, for her constant help in numerous department and university formalities essential for completion of this work.
# Contents

Abstract .................................................. ii
Acknowledgements ........................................ iv
List of Tables ............................................. vii
List of Figures ............................................ viii

1 A Glimpse of Fisher Information Matrix ............... 1
   1.1 Motivating Factors .................................. 2
   1.2 Fisher Information Matrix: Definition and Notation .... 4
   1.3 Estimation of Fisher Information Matrix ................. 5
      1.3.1 Current Re-sampling Algorithm: No use of Prior Information 6

2 Improved Re-sampling Algorithm ....................... 14
   2.1 Additional Notation .................................. 15
   2.2 Modified Re-sampling Algorithm using Prior Information . 16
   2.3 Theoretical Basis for the Modified Re-sampling Algorithm .... 20
      2.3.1 Case 1: only the measurements of \( \ell \) are available 20
      2.3.2 Case 2: measurements of \( g \) are available ............... 37

3 Numerical Illustrations and Discussions ............... 46
   3.1 Example 1 ........................................... 46
   3.2 Example 2 ........................................... 48
   3.3 Discussions on results ................................ 52

4 Conclusions and Future Research ....................... 57
   4.1 Summary of Contributions Made ........................ 58
   4.2 Suggestions for Future Research ....................... 58

Bibliography .............................................. 60
List of Tables

3.1 Example 1 – MSE and MSE reduction of FIM estimates based on $N = 2000$ pseudo data (results on variance are reported within parentheses). 

3.2 Example 1 – MSE comparison for $\hat{F}_n$ and $\tilde{F}_n$ for only the unknown elements of $F_n(\theta)$ according to $F_n^{(\text{given})}$ based on $N = 100000$ pseudo data (similar results on variance are reported within parentheses, $A = \sum_{i=1}^{p} \sum_{j \in I_i} \text{var}[\hat{F}_{ij}|\theta]$ and $B = \sum_{i=1}^{p} \sum_{j \in I_i} \text{var}[\tilde{F}_{ij}|\theta]$). 

3.3 Example 2 – MSE and MSE reduction of FIM estimates based on $N = 40000$ pseudo data (results on variance are reported within parentheses). 

3.4 Example 2 – MSE comparison for $\hat{F}_n$ and $\tilde{F}_n$ for only the unknown elements of $F_n(\theta)$ according to $F_n^{(\text{given})}$ based on $N = 40000$ pseudo data (similar results on variance are reported within parentheses $A = \sum_{i=1}^{p} \sum_{j \in I_i} \text{var}[\hat{F}_{ij}|\theta]$ and $B = \sum_{i=1}^{p} \sum_{j \in I_i} \text{var}[\tilde{F}_{ij}|\theta]$). 

3.5 Example 1' – Variance comparison between theoretical results (correct up to a bias term as discussed in Section 2.3) and simulation results (with $N = 2000$).
List of Figures

1.1 Fisher information matrix with known elements as marked; void part consists of unknown elements. 

1.2 Schematic model for forming FIM estimate, $\hat{F}_{M,N}$; adapted from [13].

2.1 Schematic of algorithm for computing the new FIM estimate, $\hat{F}_n$.

3.1 $F_{30}^{(\text{given})}$ for Example 2 with known elements as marked; void part consists of unknown elements to be estimated by using MC based resampling algorithms.
Chapter 1

A Glimpse of Fisher Information Matrix

The Fisher information matrix (FIM) plays a key role in estimation and identification [12, Section 13.3] and information theory [3, Section 17.7]. A standard problem in the practical application and theory of statistical estimation and identification is to estimate the unobservable parameters, \( \theta \), of the probability distribution function from a set of observed data set, \( \{Z_1, \cdots, Z_n\} \), drawn from that distribution [4]. The FIM is an indicator of the amount of information contained in this observed data set about the unobservable parameter, \( \theta \). If \( \hat{\theta} \) is an unbiased estimator of \( \theta^* \) representing the true value of \( \theta \), then the inverse of the FIM computed at \( \theta^* \) provides the Cramér-Rao lower bound of the covariance matrix of \( \hat{\theta} \), implying that the difference between \( \text{cov}(\hat{\theta}) \) and the inverse of the FIM computed at \( \theta^* \) is a positive semi-definite
matrix. Therefore, the name *information matrix* is used to indicate that a larger FIM (in the matrix sense – positive semi-definiteness as just mentioned) is associated with a smaller covariance matrix (*i.e.*, more information), while a smaller FIM is associated with a larger covariance matrix (*i.e.*, less information). Some important areas of applications of FIM include, to name a few, confidence interval computation of model parameter [2, 7, 16], configuration of experimental design involving linear [9] and nonlinear [12, Section 17.4] models, and determination of noninformative prior distribution (Jeffreys’ prior) for Bayesian analysis [8].

### 1.1 Motivating Factors

However, analytical determination of the FIM in a general setting, specially in nonlinear models [16], may be a formidable undertaking due to intractable modeling requirements and/or intractable high-dimensional integration. To avoid this analytical problem, a computational technique based on Monte Carlo simulation technique, called resampling approach [12, Section 13.3.5], [13], may be employed to estimate the FIM. In practice, there may be instances when some elements of the FIM are analytically known from prior information while the other elements are unknown (and need to be estimated). In a recent work [4, 5], a FIM of size $22 \times 22$ was observed to have the structure as shown in Figure 1.1.

In such cases, the above resampling approach, however, still yields the *full* FIM without taking any advantage of the information contained in the analytically known
elements. The prior information related to known elements of the FIM is not incorporated while employing this algorithm for estimation of the unknown elements. The resampling-based estimates of the known elements are also “wasted” because these estimates are simply replaced by the analytically known elements. The issue yet to be examined is whether there is a way of focusing the averaging process (required in the resampling algorithm) — on the elements of interest (unknown elements that need to be estimated) — that is more effective than simply extracting the estimates of those elements from the full FIM estimated by employing the existing resampling algorithm.

The current work presents a modified and improved (in the sense of variance reduction) version of the resampling approach for estimating the unknown elements of the FIM by “borrowing” the information contained in the analytically known elements. The final outcome exactly preserves the known elements of the FIM and
produces variance reduction for the estimators of the unknown elements that need to be estimated.

The main idea of the current work is presented in chapter 2.2 and the relevant theoretical basis is highlighted in section 2.3. The proposed resampling algorithm, that is similar in some sense to the one for Jacobian/Hessian estimates presented earlier [14], modifies and improves the current resampling algorithm by simultaneously preserving the known elements of the FIM and yielding better (in the sense of variance reduction) estimates of the unknown elements. The effectiveness and efficiencies of the new resampling algorithm is illustrated by numerical examples in chapter 3. Chapter 4 infers the conclusions from the current work.

1.2 Fisher Information Matrix: Definition and Notation

Consider a set of $n$ random data vector $\{Z_1, \cdots, Z_n\}$ and stack them in $Z_n$, i.e., $Z_n = [Z_1^T, \cdots, Z_n^T]^T$. Here, the superscript $T$ is transpose operator. Let the multivariate joint probability density or mass (or hybrid density/mass) function (pdf) of $Z_n$ be denoted by $p_{Z_n}(\cdot|\theta)$ that is parameterized by a vector of parameters, $\theta = [\theta_1, \cdots, \theta_p]^T \in \Theta \subseteq \mathbb{R}^p$ in which $\Theta$ is the set consisting of allowable values for $\theta$. The likelihood function of $\theta$ is then given by $\ell(\theta|Z_n) = p_{Z_n}(Z_n|\theta)$ and the associated log-likelihood function $L$ by $L(\theta|Z_n) \equiv \ln \ell(\theta|Z_n)$. 
Let us define the gradient vector \( g \) of \( L \) by \( g(\theta|Z_n) = \partial L(\theta|Z_n)/\partial \theta \) and the Hessian matrix \( H \) by \( H(\theta|Z_n) = \partial^2 L(\theta|Z_n)/\partial \theta \partial \theta^T \). Here, \( \partial(\cdot)/\partial \theta \) is the \( p \times 1 \) column vector of elements \( \partial(\cdot)/\partial \theta_i, i = 1, \cdots, p \), \( \partial(\cdot)/\partial \theta^T = [\partial(\cdot)/\partial \theta]^T \), and \( \partial^2(\cdot)/\partial \theta \partial \theta^T \) is the \( p \times p \) matrix of elements \( \partial^2(\cdot)/\partial \theta_i \partial \theta_j, i, j = 1, \cdots, p \). Then, the \( p \times p \) FIM, \( F_n(\theta) \), is defined [12, Section 13.3.2] as follows,

\[
F_n(\theta) \equiv E \left[ g(\theta|Z_n) \cdot g^T(\theta|Z_n) \right] = -E \left[ H(\theta|Z_n) \right], \tag{1.1}
\]

provided that the derivatives and expectation (the expectation operator \( E \) is with respect to the probability measure of \( Z_n \)) exist. The equality ‘\( = \)’ in (1.1) is followed [1, Proposition 3.4.4], [12, p. 352 – 353] by assuming that \( L \) is twice differentiable with respect to \( \theta \) and the regularity conditions [1, Section 3.4.2] hold for the likelihood function, \( \ell \). The Hessian-based form above is more amenable to the practical computation for FIM than the gradient-based form that is used for defining the FIM. It must be remarked here that there might be cases when the observed FIM, \( -H(\theta|Z_n) \equiv -\partial^2 L(\theta|Z_n)/\partial \theta \partial \theta^T \), at a given data set, \( Z_n \), is preferable to the actual FIM, \( F_n(\theta) = -E[H(\theta|Z_n)|\theta] \), from an inference point of view and/or computational simplicity (see e.g., [6] for scalar case, \( \theta = \theta_1 \)).

### 1.3 Estimation of Fisher Information Matrix

The analytical computation of \( F_n(\theta) \), in a general case, is a daunting task since it involves evaluation of the first or second order derivative of \( L \) and expectation of
the resulting multivariate, often non-linear, function that become difficult or next to impossible. This analytical problem can be circumvented by employing a recently introduced simulation-based technique, resampling approach [12, Section 13.3.5], [13]. This algorithm is essentially based on producing a set of large (say $N$) number of Hessian estimates from either the values of the log-likelihood function or (if available) its exact stochastic gradient that, in turn, are computed from a set of pseudo data vector, \{\text{\textbf{Z}}_{\text{pseudo}}(1), \cdots, \text{\textbf{Z}}_{\text{pseudo}}(N)\}, with each \text{\textbf{Z}}_{\text{pseudo}}(i), i = 1, \cdots, N, being digitally simulated from $p_{\text{\textbf{Z}}_n} (\cdot | \theta)$ and statistically independent of each other. The set of pseudo data vector acts as a proxy for the observed data set in the resampling algorithm. The average of the negative of these Hessian estimates is reported as an estimate of the $F_n(\theta)$.

A brief outline of the existing resampling algorithm is provided in the following subsection.

1.3.1 Current Re-sampling Algorithm: No use of Prior Information

For $i$-th pseudo data, let the $k$-th estimate of the Hessian matrix, $\text{\textbf{H}}(\theta|\text{\textbf{Z}}_{\text{pseudo}}(i))$, in the resampling algorithm, be denoted by $\hat{\text{\textbf{H}}}^{(i)}_k$. Then, $\hat{\text{\textbf{H}}}^{(i)}_k$, as per resampling scheme, is computed as [13],

$$
\hat{\text{\textbf{H}}}^{(i)}_k = \frac{1}{2} \left\{ \frac{\delta \textbf{G}^{(i)}_k}{2c} \left[ \Delta_{k1}^{-1}, \cdots, \Delta_{kp}^{-1} \right] + \left( \frac{\delta \textbf{G}^{(i)}_k}{2c} \left[ \Delta_{k1}^{-1}, \cdots, \Delta_{kp}^{-1} \right] \right)^T \right\}, \quad (1.2)
$$
in which $c > 0$ is a small number, $\delta G_k^{(i)} \equiv G(\theta + c\Delta_k|\text{Z}_{\text{pseudo}}(i)) - G(\theta - c\Delta_k|\text{Z}_{\text{pseudo}}(i))$

and the perturbation vector $\Delta_k = [\Delta_{k1}, \cdots, \Delta_{kp}]^T$ is a user-generated random vector statistically independent of $\text{Z}_{\text{pseudo}}(i)$. The random variables, $\Delta_{k1}, \cdots, \Delta_{kp}$, are mean-zero and statistically independent and, also the inverse moments, $E[1/\Delta_{km}]$, $m = 1, \cdots, p$, are finite.

The symmetrizing operation (the multiplier $1/2$ and the indicated sum) as shown in (1.2) is useful in optimization problems to compute a symmetric estimate of the Hessian matrix with finite samples [11]. This also maintains a symmetric estimate of $\mathbf{F}_n(\theta)$, which itself is a symmetric matrix.

Depending on the setting, $\mathbf{G}(\cdot|\text{Z}_{\text{pseudo}}(i))$, as required in $\delta G_k^{(i)}$, represents the $k$-th direct measurement or approximation of the gradient vector $\mathbf{g}(\cdot|\text{Z}_{\text{pseudo}}(i))$. If the direct measurement or computation of the exact gradient vector $\mathbf{g}$ is feasible, $\mathbf{G}(\theta \pm c\Delta_k|\text{Z}_{\text{pseudo}}(i))$ represent the direct $k$-th measurements of $\mathbf{g}(\cdot|\text{Z}_{\text{pseudo}}(i))$ at $\theta \pm c\Delta_k$.

If the direct measurement or computation of $\mathbf{g}$ is not feasible, $\mathbf{G}(\theta \pm c\Delta_k|\text{Z}_{\text{pseudo}}(i))$ represents the $k$-th approximation of $\mathbf{g}(\theta \pm c\Delta_k|\text{Z}_{\text{pseudo}}(i))$ based on the values of $L(\cdot|\text{Z}_{\text{pseudo}}(i))$.

If the direct measurements or computations of $\mathbf{g}$ are not feasible, $\mathbf{G}$ in (1.2) can be computed by using the classical finite-difference (FD) technique [12, Section 6.3] or the simultaneous perturbation (SP) gradient approximation technique [10], [12, Section 7.2] from the values of $L(\cdot|\text{Z}_{\text{pseudo}}(i))$. For the computation of gradient approximation based on the values of $L$, there are advantages to using one-sided [12,
p. 199] SP gradient approximation (relative to the standard two-sided SP gradient approximation) in order to reduce the total number of function measurements or evaluations for $L$. The SP technique for gradient approximation is quite useful when $p$ is large and usually superior to FD technique for estimation of $F_n(\theta)$ when the resampling approach is employed to estimate the FIM. The formula for the one-sided gradient approximation using SP technique is given by,

$$G^{(1)}(\theta \pm c\Delta_k|Z_{\text{pseudo}}(i)) = \left(1/\tilde{c}\right) \left[ L(\theta + \tilde{c}\Delta_k \pm c\Delta_k|Z_{\text{pseudo}}(i)) - L(\theta \pm c\Delta_k|Z_{\text{pseudo}}(i)) \right] \begin{bmatrix} \tilde{\Delta}^{-1}_{k1} \\ \vdots \\ \tilde{\Delta}^{-1}_{kp} \end{bmatrix}, \quad (1.3)$$

in which superscript (1) in $G^{(1)}$ indicates that it is one-sided gradient approximation ($G = G^{(1)}$), $\tilde{c} > 0$ is a small number and $\tilde{\Delta}_k = [\tilde{\Delta}_k 1, \cdots, \tilde{\Delta}_k p]^T$ is generated in the same statistical manner as $\Delta_k$, but otherwise statistically independent of $\Delta_k$ and $Z_{\text{pseudo}}(i)$. It is usually recommended that $\tilde{c} > c$. Note that $G^{(1)}(\cdot)$ in (1.3) not only depends on $\theta$, $c$, $\Delta_k$ and $Z_{\text{pseudo}}(i)$ but also depends on $\tilde{c}$ and $\tilde{\Delta}_k$ that, however, have been suppressed for notational clarity.

At this stage, let us also formally state that the perturbation vectors, $\Delta_k$ and $\tilde{\Delta}_k$, satisfy the following condition [10, 11], [12, Chapter 7].

C.1: (Statistical properties of the perturbation vector) The random variables, $\Delta_{km}$ (and $\tilde{\Delta}_{km}$), $k = 1, \cdots, N$, $m = 1, \cdots, p$, are statistically independent and almost surely (a.s.) uniformly bounded for all $k$, $m$, and, are also
mean-zero and symmetrically distributed satisfying \( E[|1/\Delta_{km}|] < \infty \) (and \( E[|1/\tilde{\Delta}_{km}|] < \infty \)).

If the pseudo data vectors are expensive to simulate relative to the Hessian estimate, then it is preferred to generate several Hessian estimates by generating more than one (say, \( M > 1 \)) perturbation vectors, \( \Delta_1, \ldots, \Delta_M \), for each pseudo data vector, \( Z_{\text{pseudo}}(i) \), and, thus, computing \( M \) Hessian estimates, \( \hat{H}_1^{(i)}, \ldots, \hat{H}_M^{(i)} \), by employing (1.2) using these \( M \) perturbation vectors and the pseudo data vector, \( Z_{\text{pseudo}}(i) \). Let the average of, say, \( k \) Hessian estimates, \( \bar{H}_1^{(i)}, \ldots, \bar{H}_k^{(i)} \), for the \( i \)-th pseudo data, \( Z_{\text{pseudo}}(i) \), be denoted by \( \bar{H}_k^{(i)} \). The computation of \( \bar{H}_k^{(i)} \) can be conveniently carried out by using the standard recursive representation of sample mean in contrast to storing the matrices and averaging later. The recursive representation of \( \bar{H}_k^{(i)} \) is given by,

\[
\bar{H}_k^{(i)} = \frac{k-1}{k} \bar{H}_{k-1}^{(i)} + \frac{1}{k} \hat{H}_k^{(i)}, \quad k = 1, \ldots, M, \tag{1.4}
\]

in which at the first step \( (k = 1) \), \( \bar{H}_0^{(i)} \) in the right-hand-side can be set to a matrix of zeros or any arbitrary matrix of finite-valued elements to obtain the correct result. If the sample mean of total \( M \) Hessian estimates is now denoted by \( \bar{H}^{(i)} \equiv \bar{H}_M^{(i)} \) (for the sake of simplification of notation), then an estimate, \( \bar{F}_{M,N} \), of \( F_n(\theta) \) is computed by averaging \( \bar{H}^{(1)}, \ldots, \bar{H}^{(N)} \) and, taking the negative value of the resulting average. A recursive representation identical to (1.4) for computing the sample mean of \( \bar{H}^{(1)}, \ldots, \bar{H}^{(N)} \) is also recommended to avoid storage of the matrices. The current resampling algorithm is schematically shown in Figure 1.2.
However, it should be noted that $M = 1$ has certain optimality properties [13] and it is assumed throughout the work presented in this thesis that for each pseudo data vector, only one perturbation vector is generated and, thus, only one Hessian estimate is computed. The current work can, however, be readily extended to the case when $M > 1$. Therefore, from now on, the index of the pseudo data vector will be changed from $i$ to $k$. Consequently, the pseudo data vector will be denoted by $Z_{\text{pseudo}}(k), k = 1, \cdots, N$, the difference in gradient and the Hessian estimate in (1.2), respectively, will simply be denoted by $\delta G_k$ and $\hat{H}_k, k = 1, \cdots, N$, and the notation of the one-sided gradient approximation in (1.3) will take the form of $G^{(1)}(\theta \pm c\Delta_k|Z_{\text{pseudo}}(k))$. Finally, the following simplification of notation for the estimate of the FIM will also be used from now on,

$$\hat{F}_n \equiv \hat{F}_{1,N}. \quad (1.5)$$
The set of simplified notations as just mentioned essentially conforms to the work in [12, Section 13.3.5] that is a simplified version of [13] in which the resampling approach was rigorously presented.

Let us also assume that the moments of $\Delta_{km}$ and $1/\Delta_{km}$ (and, of $\tilde{\Delta}_{km}$ and $1/\tilde{\Delta}_{km}$) up to fifth order exist (this condition will be used later in Section 2.3. Since $\Delta_{km}$ (and $\tilde{\Delta}_{km}$) is symmetrically distributed, $1/\Delta_{km}$ (and $1/\tilde{\Delta}_{km}$) is also symmetrically distributed implying that

I: (Statistical properties implied by C.1) All the odd moments of $\Delta_{km}$ and $1/\Delta_{km}$ (and of $\tilde{\Delta}_{km}$ and $1/\tilde{\Delta}_{km}$) up to 5-th order are zeros, $E[(\Delta_{km})^q] = 0$ and $E[(1/\Delta_{km})^q] = 0$ ($E[(\tilde{\Delta}_{km})^q] = 0$ and $E[(1/\tilde{\Delta}_{km})^q] = 0$), $q = 1, 3, 5$.

The random vectors $\Delta_k$ (and $\tilde{\Delta}_k$) are also independent across $k$. The random variables $\Delta_{k1}, \ldots, \Delta_{kp}$ (and $\tilde{\Delta}_{k1}, \ldots, \tilde{\Delta}_{kp}$) can also be chosen identically distributed. In fact, independent and identically distributed (i.i.d.) (across both $k$ and $m$) mean-zero random variable satisfying C.1 is a perfectly valid choice for $\Delta_{km}$ (and $\tilde{\Delta}_{km}$). In particular, Bernoulli $\pm 1$ random variable for $\Delta_{km}$ (and $\tilde{\Delta}_{km}$) is a valid — but not the necessary — choice among other probability distributions satisfying C.1.

It must be noted from (1.2) and (1.3) that the current resampling algorithm would will satisfactorily provided the following condition is satisfied.

C.2: (Smoothness of $L$) The log-likelihood function, $L$, is twice continuously differentiable with respect to $\theta$ and all the possible second order derivatives are a.s. (a.s. with respect to the probability measure of $Z_{\text{pseudo}}(k)$) continuous.
and uniformly (in \( k \)) bounded \( \forall \theta \in S_\theta \subset \Theta \subseteq \mathbb{R}^p \) in which \( S_\theta \) is a small region containing \( \hat{\theta} \) that represents the value of \( \theta \) at which the FIM needs to estimated. (For example, if \( |\Delta_{kl}| \leq \rho \) and \( |	ilde{\Delta}_{kl}| \leq \tilde{\rho}, \forall k, l \), for some small numbers \( \rho, \tilde{\rho} > 0 \), then it is sufficient to consider \( S_\theta \) as a hypercube with length of its side being \( 2(c\rho + \tilde{c}\tilde{\rho}) \) and center being \( \hat{\theta} \) in \( \mathbb{R}^p \).

In all the gradient approximation techniques as mentioned earlier (FD technique, two-sided SP technique and one-sided SP technique), the gradient approximation, \( G(\theta|Z_{\text{pseudo}}(k)) \), has \( \mathcal{O}_Z(\tilde{c}^2) \) bias provided the following is satisfied.

C.2' (Smoothness of \( L \)) The log-likelihood function, \( L \), is thrice continuously differentiable with respect to \( \theta \) and all the possible third order derivatives are a.s. continuous and bounded \( \forall \theta \in S_\theta \).

The exact form of the bias terms are, however, different for different techniques [12, Section 6.4.1, 7.2.2]. The subscripts in \( \mathcal{O}_Z(\cdot) \) explicitly indicate that the ‘big-O’ term depends on \( Z_{\text{pseudo}}(k) \). This term is such that \( |\mathcal{O}_Z(\tilde{c}^2)/\tilde{c}^2| < \infty \) a.s. as \( \tilde{c} \rightarrow 0 \). The explicit expression of bias, when one-sided SP technique is employed, is reported in Section 2.3.1. In the case where the direct measurements or computations of \( g \) are feasible, \( G \) is an unbiased gradient estimate (measurement) of \( g \).

It is shown in the next chapter that the FIM estimate in (1.5) is accurate within an \( \mathcal{O}(c^2) + \mathcal{O}(\tilde{c}^2) \) bias in the case where only the measurements of the log-likelihood function \( L \) are available and within an \( \mathcal{O}(c^2) \) bias in the case where the measurements
of the unbiased exact stochastic gradient vector $\mathbf{g}$ are available provided the following condition (stronger than C.2 and C.2') is satisfied.

C.3: **(Smoothness of $L$)** The log-likelihood function, $L$, is four times continuously differentiable with respect to $\theta$ and all the possible fourth order derivatives are a.s. continuous and uniformly (in $k$) bounded $\forall \theta \in S_\theta$.

Here, the ‘big-$O$’ terms, $O(c^2)$ and $O(\tilde{c}^2)$, satisfy the conditions, $|O(c^2)/c^2| < \infty$ as $c \longrightarrow 0$ and $|O(\tilde{c}^2)/\tilde{c}^2| < \infty$ as $\tilde{c} \longrightarrow 0$. The FIM estimate can be made as accurate as desired through reducing $c$ and $\tilde{c}$ and increasing the number of $\hat{H}_k$ values being averaged.

In the next chapter, a modification is proposed in the existing resampling algorithm that would enhance the statistical characteristics of the estimator of $F_n(\theta)$, a part of which is analytically known a priori.
Chapter 2

Improved Re-sampling Algorithm

Let the \( k \)-th estimate of Hessian matrix, \( \mathbf{H}(\theta|Z_{\text{pseudo}}(k)) \), per proposed resampling algorithm be denoted by \( \tilde{\mathbf{H}}_k \). In this chapter, the estimator, \( \tilde{\mathbf{H}}_k \), is shown separately for two different cases: Case 1 - when only the measurements of the log-likelihood function \( L \) are available and, Case 2: when measurements of the exact gradient vector \( \mathbf{g} \) are available. To contrast the two cases, the superscript \( (L) \) is used in \( \hat{\mathbf{H}}_k^{(L)} \) and \( \hat{\mathbf{H}}_k^{(L)} \) to represent the dependence of \( \tilde{\mathbf{H}}_k \) and \( \hat{\mathbf{H}}_k \) on \( L \) measurements for Case 1 and, similarly, the superscript \( (g) \) in \( \hat{\mathbf{H}}_k^{(g)} \) and \( \hat{\mathbf{H}}_k^{(g)} \) for Case 2. (The superscripts, \( (L) \) or \( (g) \), should not be confused with the superscript, \( (i) \), as appeared in (1.2) which referred to the pseudo data vector. The superscript \( (i) \) was eventually suppressed in consideration of \( M \) being 1 as indicated at p. 10.)
2.1 Additional Notation

Denote the \((i, j)\)-th element of \(F_n(\theta)\) by \(F_{ij}(\theta)\) (non-bold character and suppressing the subscript, \(n\), in the symbolic notation representing the element of \(F_n(\theta)\)) for simplification of notation. Let \(I_i, i = 1, \cdots, p\), be the set of column indices of the known elements of \(F_n(\theta)\) for the \(i\)-th row and \(I_i^c\) be the complement of \(I_i\). Consider a \(p \times p\) matrix \(F_{ij}^{(\text{given})}\) whose \((i, j)\)-th element, \(F_{ij}^{(\text{given})}\), is defined as follows,

\[
F_{ij}^{(\text{given})} = \begin{cases} 
F_{ij}(\theta), & \text{if } j \in I_i, \\
0, & \text{otherwise}
\end{cases}, \quad i = 1, \cdots, p. \quad (2.1)
\]

Consider another \(p \times p\) matrix \(D_k\) defined by,

\[
D_k = \begin{bmatrix}
\Delta_{k1} & \Delta_{k1}^{-1} & \cdots & \Delta_{k1}^{-1} \\
\Delta_{k2} & \Delta_{k2}^{-1} & \cdots & \Delta_{k2}^{-1} \\
\vdots & \vdots & \ddots & \vdots \\
\Delta_{kp} & \Delta_{kp}^{-1} & \cdots & \Delta_{kp}^{-1}
\end{bmatrix}
= \Delta_k \begin{bmatrix}
\Delta_{k1}^{-1} \\
\vdots \\
\Delta_{kp}^{-1}
\end{bmatrix}, \quad (2.2)
\]

\(D_k\) together with a corresponding matrix \(D_k\) based on replacing all \(\Delta_{ki}\) in \(D_k\) by the corresponding \(\tilde{\Delta}_{ki}\) (note that \(D_k\) is symmetric when the perturbations are i.i.d. Bernoulli distributed).
2.2 Modified Re-sampling Algorithm using Prior Information

The new estimate, $\tilde{H}_k$, is extracted from $\tilde{H}_{k0}$ that is defined next separately for Case 1 and Case 2:

Case 1: only the measurements of the log-likelihood function $L$ are available,

$$\tilde{H}^{(L)}_{k0} = \hat{H}^{(L)}_k - \frac{1}{2} \left[ \mathcal{D}^T_k \left( -F^{\text{given}}_n \right) \mathcal{D}_k + \left( \mathcal{D}^T_k \left( -F^{\text{given}}_n \right) \mathcal{D}_k \right)^T \right]. \quad (2.3)$$

Case 2: measurements of the exact gradient vector $g$ are available,

$$\tilde{H}^{(g)}_{k0} = \hat{H}^{(g)}_k - \frac{1}{2} \left[ \left( -F^{\text{given}}_n \right) \mathcal{D}_k + \left( -F^{\text{given}}_n \right) \mathcal{D}_k \right]^T \right]. \quad (2.4)$$

The estimates, $\tilde{H}^{(L)}_k$ and $\tilde{H}^{(g)}_k$, are readily obtained from, respectively, $\tilde{H}^{(L)}_{k0}$ in (2.3) and $\tilde{H}^{(g)}_{k0}$ in (2.4) by replacing the $(i, j)$-th element of $\tilde{H}^{(L)}_{k0}$ and $\tilde{H}^{(g)}_{k0}$ with known values of $-F_{ij}(\theta)$, $j \in I_i$, $i = 1, \cdots, p$. The new estimate, $\tilde{F}_n$, of $F_n(\theta)$ is then computed by averaging the Hessian estimates, $\tilde{H}_k$, and taking negative value of the resulting average. For convenience and, also since the main objective is to estimate the FIM, not the Hessian matrix, the replacement of the $(i, j)$-th element of $\tilde{H}_{k0}$ with known values of $-F_{ij}(\theta)$, $j \in I_i$, $i = 1, \cdots, p$, is not required at each $k$-th step. The matrix, $\tilde{F}_{n0}$, can be first obtained by computing the (negative) average of the matrices, $\tilde{H}_{k0}$, and subsequently, the $(i, j)$-th element of $\tilde{F}_{n0}$ can be replaced with the analytically known elements, $F_{ij}(\theta)$, $j \in I_i$, $i = 1, \cdots, p$, of $F_n(\theta)$ to yield the new estimate, $\tilde{F}_n$. Consequently, replacing needs to be done only once at the end of averaging. However, if a better estimate of Hessian matrix is also required (e.g., in optimization [11, 14]),
the replacing as mentioned above for Hessian estimates is recommended at each \( k \)-th step.

The matrices, \( \hat{H}_k^{(L)} \) in (2.3) or \( \hat{H}_k^{(g)} \) in (2.4), need to be computed by using the existing resampling algorithm implying that \( \hat{H}_k^{(L)} \equiv \hat{H}_k \) or \( \hat{H}_k^{(g)} \equiv \hat{H}_k \) as appropriate with \( \hat{H}_k \) being given by (1.2). Note that \( F_n^{(\text{given})} \) as shown in the right-hand-sides of (2.3) and (2.4) is known by (2.1). It must be noted as well that the random perturbation vectors, \( \Delta_k \) in \( D_k \) and \( \tilde{\Delta}_k \) in \( \tilde{D}_k \), as required in (2.3) and (2.4) must be the same simulated values of \( \Delta_k \) and \( \tilde{\Delta}_k \) used in the existing resampling algorithm while computing the \( k \)-th estimate, \( \hat{H}_k^{(L)} \) or \( \hat{H}_k^{(g)} \).

Next a summary of the salient steps, required to produce the estimate \( \tilde{F}_n \) \( i.e. \), with appropriate superscript, \( \tilde{F}_n^{(L)} \) or \( \tilde{F}_n^{(g)} \) of \( F_n(\theta) \) per modified resampling algorithm as proposed here, is presented below. Figure 2.1 is a schematic of the following steps.

**Step 0. Initialization:** Construct the matrix \( F_n^{(\text{given})} \) as defined by (2.1) based on the analytically known elements of FIM. Determine \( \theta \), the sample size \( n \) and the number \( (N) \) of pseudo data vectors that will be generated. Determine whether log-likelihood \( L(\cdot) \) or gradient vector \( g(\cdot) \) will be used to compute the Hessian estimates \( \hat{H}_k \). Pick a small number \( c \) (perhaps \( c = 0.0001 \)) to be used for Hessian estimation (see (1.2)) and if required another small number \( \tilde{c} \) (perhaps \( \tilde{c} = 0.00011 \)) for gradient approximation (see (1.3)). Set \( k = 1 \).
Step 1. At the $k$-th step perform the following tasks,

a. **Generation of pseudo data:** Based on $\theta$, generate by MC simulation technique the $k$-th pseudo data vector of $n$ pseudo measurements $Z_{\text{pseudo}}(k)$.

b. **Computation of $\hat{H}_k$:** Generate the $k$-th perturbation vector $\Delta_k$ by satisfying $C.1$ and if required the other perturbation vector $\tilde{\Delta}_k$ by satisfying $\tilde{C}.1$ for gradient approximation. Using the $k$-th pseudo data vector $Z_{\text{pseudo}}(k)$ and the perturbation vector(s) $\Delta_k$ or/and $\tilde{\Delta}_k$, evaluate $\hat{H}_k$ (i.e., $\hat{H}_k^{(L)}$ or $\hat{H}_k^{(g)}$) by using (1.2).

c. **Computation of $D_k$ and $\tilde{D}_k$:** Use $\Delta_k$ or/and $\tilde{\Delta}_k$, as generated in the above step, to construct the matrices, $D_k$ or/and $\tilde{D}_k$, as defined in section 2.1.

d. **Computation of $\tilde{H}_{k0}$:** Modify $\hat{H}_k$ as produced in **Step 1b** by employing (2.3) or (2.4) as appropriate in order to generate $\tilde{H}_{k0}$ (i.e., $\tilde{H}_{k0}^{(L)}$ or $\tilde{H}_{k0}^{(g)}$).

Step 2. **Average of $\tilde{H}_{k0}$:** Repeat **Step 1** until $N$ estimates, $\tilde{H}_{k0}$, are produced. Compute the (negative) mean of these $N$ estimates. (The standard recursive representation of sample mean can be used here to avoid the storage of $N$ matrices, $\tilde{H}_{k0}$). The resulting (negative) mean is $\tilde{F}_{n0}$.

Step 3. **Evaluation of $\tilde{F}_n$:** The new estimate, $\tilde{F}_n$, of $F_n(\theta)$ per modified re-
sampling algorithm is simply obtained by replacing the \((i, j)\)-th element of \(\tilde{F}_{n0}\) with the analytically known elements, \(F_{ij}(\theta), j \in \mathbb{I}_i, i = 1, \cdots, p\), of \(F_n(\theta)\). To avoid the possibility of having a non-positive semi-definite estimate, it may be desirable to take the symmetric square root of the square of the estimate (the `sqrtm` function in MATLAB may be useful here).

![Schematic of algorithm for computing the new FIM estimate, \(\tilde{F}_n\).](image)

Figure 2.1: Schematic of algorithm for computing the new FIM estimate, \(\tilde{F}_n\).

The new estimator, \(\hat{F}_n\), is better than \(\check{F}_n\) in the sense that it would preserve exactly the analytically known elements of \(F_n(\theta)\) as well as reduce the variances of the estimators of the unknown elements of \(F_n(\theta)\).

The next section sketches the theoretical basis of the modified resampling algorithm as proposed in this section.
2.3 Theoretical Basis for the Modified Re-sampling Algorithm

It should be remarked here for the sake of clarification that the dependence of $\tilde{H}_k$ and $\hat{H}_k$ on $\theta$, $Z_{\text{pseudo}}(k)$, $\Delta_k$, $c$ or/and $\tilde{\Delta}_k$ and $\tilde{c}$ are suppressed throughout this work for simplification of notation. Now, for further notational simplification, the subscript ‘pseudo’ in $Z_{\text{pseudo}}(k)$ and the dependence of $Z(k)$ on $k$ would be suppressed (note that $Z_{\text{pseudo}}(k)$ is identically distributed across $k$ because $Z_{\text{pseudo}}(k) \sim p_{Z_n}(\cdot|\theta)$). Since, $\Delta_k$ is usually assumed to be statistically independent across $k$ and an identical condition for $\tilde{\Delta}_k$ is also assumed, the dependence on $k$ would also be suppressed in the future development.

Let the $(i, j)$-th element of $\hat{H}_k$ and $\tilde{H}_k$ be, respectively, denoted by $\hat{H}_{ij}$ and $\tilde{H}_{ij}$ with the appropriate superscript. In the following, the mean and variance of both $\hat{H}_{ij}$ and $\tilde{H}_{ij}$ are computed for two cases in two different subsections. It is also shown that the variance of the new estimator, $\tilde{H}_{ij}$, is less than the variance of the existing estimator, $\hat{H}_{ij}$, yielding a better estimate of $F_n(\theta)$ in the process.

2.3.1 Case 1: only the measurements of $L$ are available

The mean and variance of $\hat{H}_{ij}^{(L)}$ are computed first followed by the mean and variance of $\tilde{H}_{ij}^{(L)}$. The subsequent comparison of variances will show the superiority of $\tilde{H}_k^{(L)}$, which leads to the superiority of $\tilde{F}_n^{(L)}$. 

20
In deducing the expressions of mean and variance in the following, the conditions, C.1 and C.3, are required along with the following condition as stated below. (The condition, C.2, only ensures that employing the resampling algorithm would yield some estimate but it cannot guarantee the boundedness of several ‘big-O’ terms as mentioned earlier or as will be encountered while deducing the expressions of mean and variance below.)

C.4: **(Existence of covariance involving \( H_{lm}, H_{lm,s} \) and \( H_{lm,rs} \))** All the combinations of covariance terms involving \( H_{lm}(\theta|Z) \), \( H_{lm,s}(\theta|Z) \) and \( H_{lm,rs}(\theta|Z) \), \( l, m, s, r = 1, \cdots, p \), exist \( \forall \theta \in S_\theta \) (see C.2 for definition of \( S_\theta \)).

**Mean and Variance of \( \hat{H}_{ij}^{(L)} \)**

It is assumed here that the gradient estimate is based on one-sided gradient approximation using SP technique given by (1.3). Based on a Taylor expansion, the \( i \)-th component of \( G^{(1)}(\theta|Z) \), \( i = 1, \cdots, p \), that is an approximation of the \( i \)-th component, \( g_i(\theta|Z) \equiv \partial L(\theta|Z)/\partial \theta_i \), of \( g(\theta|Z) \) based on the values of \( L(\cdot|Z) \), is given by,

\[
G^{(1)}_i(\theta|Z) = \frac{L(\theta + \hat{c} \Delta|Z) - L(\theta|Z)}{\hat{c} \Delta_i} \tag{2.5}
\]

\[
= \frac{1}{\hat{c} \Delta_i} \left[ \left( L(\theta|Z) + \hat{c} \sum_{l=1}^{p} g_l(\theta|Z) \Delta_l + \frac{1}{2} \hat{c}^2 \sum_{l=1}^{p} \sum_{m=1}^{p} H_{lm}(\theta|Z) \Delta_m \Delta_l \right)
+ \frac{\hat{c}^3}{6} \sum_{l=1}^{p} \sum_{m=1}^{p} \sum_{s=1}^{p} \frac{\partial^3 L(\theta|Z)}{\partial \theta_s \partial \theta_m \partial \theta_l} \Delta_s \Delta_m \Delta_l \right] - L(\theta|Z)
\]

\[
= \sum_l g_l(\theta) \frac{\Delta_l}{\Delta_i} + \frac{1}{2} \hat{c} \sum_{l,m} H_{lm}(\theta) \frac{\Delta_m \Delta_l}{\Delta_i} + \frac{\hat{c}^2}{6} \sum_{l,m,s} \frac{\partial H_{lm}(\theta)}{\partial \theta_s} \frac{\Delta_s \Delta_m \Delta_l}{\Delta_i}, \tag{2.6}
\]
in which \( H_{lm}(\theta|Z) \equiv \partial^2 L(\theta|Z)/\partial \theta_l \partial \theta_m \) is the \((l,m)\)-th element of \( H(\theta|Z) \), \( \overline{\theta} = \lambda(\theta + \tilde{c}\tilde{\Delta}) + (1 - \lambda)\theta = \theta + \tilde{c}\lambda\tilde{\Delta} \) (with \( \lambda \in [0,1] \) being some real number) denotes a point on the line segment between \( \theta \) and \( \theta + \tilde{c}\tilde{\Delta} \) and, in the expression after the last equality, the condition on \( Z \) is suppressed for notational clarity and, also the summations are expressed in abbreviated format where the indices span their respective and appropriate ranges. This convention would be followed throughout this work unless the explicit dependence of \( Z \) and an expanded summation format need to be invoked for the sake of clarity of the context to be discussed.

Now note that

\[
E[G_i^{(1)}(\theta|Z)|\theta, Z] = \sum_l g_l(\theta|Z)E\left[\frac{\dot{\Delta}_i}{\Delta_i}\right] + 0 + \frac{1}{6}\tilde{c}^2 \sum_{l,m,s} \frac{\partial H_{lm}(\bar{\theta}|Z)}{\partial \theta_s} E\left[\frac{\dot{\Delta}_s\dot{\Delta}_m\dot{\Delta}_l}{\dot{\Delta}_i}\right]
\]

\[
= g_i(\theta|Z) + O_Z(\tilde{c}^2)
\]

(2.7)

in which the point of evaluation, \( \theta \), is suppressed in the random ‘big-O’ term, \( O_Z(\tilde{c}^2) = (1/6)\tilde{c}^2 \sum_{l,m,s} \left[ \frac{\partial H_{lm}(\bar{\theta}|Z)}{\partial \theta_s} E[(\dot{\Delta}_s\dot{\Delta}_m\dot{\Delta}_l)/\dot{\Delta}_i] \right] \) for notational clarity. The subscript in \( O_Z(\cdot) \) explicitly indicate that this term depends on \( Z \) satisfying \( |O_Z(\tilde{c}^2)/\tilde{c}^2| < \infty \) almost surely (a.s.) (a.s. with respect to the joint probability measure of \( Z \)) as \( \tilde{c} \to 0 \) by C.1 and C.2'. The first equality follows from the fact that the second term in (2.6) vanishes upon expectation since it involves either \( E[\dot{\Delta}_r] \) or \( E[1/\dot{\Delta}_r] \), \( r = 1, \cdots, p \), both of which are zero by implication I. The second equality follows by the condition C.1.

Now \( G_i(\theta \pm c\Delta|Z) \equiv G_i^{(1)}(\theta \pm c\Delta|Z) \), that is required for computing \( \hat{H}_{ij}^{(L)} \), need to be calculated. Suppressing the condition on \( Z \) (again for simplification of no-
tation, specially when the dependence is not important in the context of the topic being discussed) and using the abbreviated summation notation as indicated earlier, \( G_i^{(1)}(\theta \pm c\Delta) \) with subsequent and further Taylor expansion can be written as follows,

\[
G_i^{(1)}(\theta \pm c\Delta) = \sum_l g_l(\theta \pm c\Delta) \frac{\tilde{\Delta}_l}{\Delta_i} + \frac{1}{2} c \sum_{l,m} H_{lm}(\theta \pm c\Delta) \frac{\tilde{\Delta}_m \tilde{\Delta}_l}{\Delta_i} + \frac{1}{6} c^2 \sum_{l,m,s} \frac{\partial H_{lm}(\theta \pm c\Delta)}{\partial \theta_s} \frac{\tilde{\Delta}_s \tilde{\Delta}_m \tilde{\Delta}_l}{\Delta_i}.
\]

(2.8)

Here, \( \theta_i^\pm \) and \( \theta_2^\pm \) denote some points on the line segments between \( \theta \) and \( \theta \pm c\Delta \), \( \overline{\theta}^\pm \) denote some points on the line segments between \( \overline{\theta} \) and \( \overline{\theta} \pm c\Delta \), the superscript, \( \pm \) sign, in \( \theta_i^\pm \), \( \theta_2^\pm \) and \( \overline{\theta}^\pm \) corresponds to whether \( \theta + c\Delta \) or \( \theta - c\Delta \) is being considered for the argument of \( G_i^{(1)}(\cdot) \) in (2.8).

Given \( G_i(\theta \pm c\Delta|Z) \equiv G_i^{(1)}(\theta \pm c\Delta|Z) \) by (2.8), the \((i, j)\)-th element of \( \hat{H}^{(L)} \) can be readily obtained from,

\[
\hat{H}^{(L)}_{ij} = \frac{1}{2} \left[ \frac{G_i(\theta + c\Delta|Z) - G_i(\theta - c\Delta|Z)}{2c\Delta_j} + \frac{G_j(\theta + c\Delta|Z) - G_j(\theta - c\Delta|Z)}{2c\Delta_i} \right].
\]

(2.9)

Consider a typical term (say, the first term in the right-hand-side of (2.9)) and denote it by \( \hat{J}^{(L)}_{ij} \) (\( J \) is to indicate Jacobian for which the symmetrizing operation should not
be used). By the use of (2.8), \( \bar{J}_{ij}^{(L)} \), then can be written as,

\[
\bar{J}_{ij}^{(L)} = \frac{G_{i}(\theta + c\Delta |Z) - G_{i}(\theta - c\Delta |Z)}{2c\Delta_{j}}
\]

\[
= \left[ \sum_{l,m} H_{lm}(\theta |Z) \frac{\Delta_{m}}{\Delta_{l}} + O_{\Delta,\Delta,Z}(c^{2}) \right] + [O_{\Delta,\Delta,Z}(\bar{c})] + [O_{\Delta,\Delta,Z}(\bar{c}^{2})]. (2.10)
\]

The subscripts in \( O_{\Delta,\Delta,Z}(\cdot) \) explicitly indicate that the random ‘big-O’ terms depend on \( \Delta, \Delta \) and \( Z \). By the use of C.1 and C.3, it should be noted that these terms are such that \( |O_{\Delta,\Delta,Z}(c^{2})/c^{2}| < \infty \) almost surely (a.s.) (a.s. with respect to the joint probability measure of \( \Delta, \Delta \) and \( Z \)) as \( c \to 0 \) and, both \( |O_{\Delta,\Delta,Z}(\bar{c})/\bar{c}| < \infty \) a.s. and \( |O_{\Delta,\Delta,Z}(\bar{c}^{2})/\bar{c}^{2}| < \infty \) a.s. as \( \bar{c} \to 0 \). The explicit expressions of the ‘big-O’ terms in (2.10) are shown below separately,

\[
O_{\Delta,\Delta,Z}(c^{2}) = \frac{1}{12}c^{2} \sum_{l,m,s,r} \left( \frac{\partial^{2}H_{lm}(\theta_{1}^{+}|Z)}{\partial \theta_{r} \partial \theta_{s}} + \frac{\partial^{2}H_{lm}(\theta_{1}^{-}|Z)}{\partial \theta_{r} \partial \theta_{s}} \right) \frac{\Delta_{r} \Delta_{s} \Delta_{m} \Delta_{l}}{\Delta_{j} \Delta_{i}},
\]

\[
O_{\Delta,\Delta,Z}(\bar{c}) = \frac{1}{4}\bar{c} \sum_{l,m,s} \left( \frac{\partial H_{lm}(\theta_{2}^{+}|Z)}{\partial \theta_{s}} + \frac{\partial H_{lm}(\theta_{2}^{-}|Z)}{\partial \theta_{s}} \right) \frac{\Delta_{s} \Delta_{m} \Delta_{l}}{\Delta_{j} \Delta_{i}},
\]

\[
O_{\Delta,\Delta,Z}(\bar{c}^{2}) = \frac{1}{12}\bar{c}^{2} \sum_{l,m,s,r} \left( \frac{\partial^{2}H_{lm}(\bar{\theta}^{+}|Z)}{\partial \theta_{r} \partial \theta_{s}} + \frac{\partial^{2}H_{lm}(\bar{\theta}^{-}|Z)}{\partial \theta_{r} \partial \theta_{s}} \right) \frac{\Delta_{r} \Delta_{s} \Delta_{m} \Delta_{l}}{\Delta_{j} \Delta_{i}}.
\]

The effects of \( O_{\Delta,\Delta,Z}(\bar{c}^{2}) \) are not included in \( O_{\Delta,\Delta,Z}(\bar{c}) \). The reason for showing \( O_{\Delta,\Delta,Z}(\bar{c}) \) separately in (2.10) is that this term vanishes upon expectation because it involves either \( E[\hat{\Delta}_{r}] \) or \( E[1/\hat{\Delta}_{r}] \), \( r = 1, \ldots, p \), both of which are zero by implication I and rest of the terms appeared in \( O_{\Delta,\Delta,Z}(\bar{c}) \) do not depend on \( \Delta \) (as mentioned earlier below (2.8) that \( \theta_{2}^{\pm} \) in \( O_{\Delta,\Delta,Z}(\bar{c}) \) depend only on \( \theta \) and \( \theta \pm c\Delta \)). The other
terms, $O\Delta_\Delta \Delta(c^2)$ and $O\Delta_\Delta \Delta(\tilde{c}^2)$, do not vanish upon expectation yielding,

\[
E[\dot{j}^{(L)}_{ij}|\theta] = E\left[\sum_{l,m} H_{lm}(\theta|Z) \frac{\Delta_m}{\Delta_j} \frac{\Delta_l}{\Delta_i} \right] + E \left[O\Delta_\Delta \Delta(c^2)|\theta\right] \\
+ E \left[O\Delta_\Delta \Delta(\tilde{c}) + E \left[O\Delta_\Delta \Delta(\tilde{c}^2)|\theta\right]\right] \\
= \sum_{l,m} E\left[H_{lm}(\theta|Z) \frac{\Delta_m}{\Delta_j} \frac{\Delta_l}{\Delta_i} \right] + O(c^2) + 0 + O(\tilde{c}^2) \\
= \sum_{l,m} E\left[H_{ij}(\theta|Z)|\theta\right] E\left[\frac{\Delta_m}{\Delta_j}\right] E\left[\frac{\Delta_l}{\Delta_i}\right] + O(c^2) + O(\tilde{c}^2) \\
= E\left[H_{ij}(\theta|Z)|\theta\right] + O(c^2) + O(\tilde{c}^2) \tag{2.11}
\]

Note that the ‘big-$O$’ terms, $O(c^2)$ and $O(\tilde{c}^2)$, satisfying $|O(c^2)/c^2| < \infty$ as $c \to 0$ and $|O(\tilde{c}^2)/\tilde{c}^2| < \infty$ as $\tilde{c} \to 0$ (by C.1 and C.3), are deterministic unlike the random ‘big-$O$’ terms in (2.10). The third equality follows from the fact that $\Delta$, $\Delta$ and $Z$ are statistically independent of each other. The fourth equality follows by the condition C.1. In (2.11), $E\left[H_{ij}(\theta|Z)|\theta\right]$ has not been substituted by $-F_{ij}(\theta)$ in order to keep the derivation of (2.11) most general since (2.11) will be applicable as well for Jacobian estimate. However, in the context of FIM, $E\left[H_{ij}(\theta|Z)|\theta\right] = -F_{ij}(\theta)$ by (Hessian-based) definition using which $E[\dot{H}^{(L)}_{ij}|\theta]$ is obtained as,

\[
E[\dot{H}^{(L)}_{ij}|\theta] = \frac{1}{2}\left(E[\dot{j}^{(L)}_{ij}|\theta] + E[\dot{j}^{(L)}_{ji}|\theta]\right) \\
= \frac{1}{2}\left(E\left[H_{ij}(\theta|Z)|\theta\right] + E\left[H_{ji}(\theta|Z)|\theta\right]\right) + O(c^2) + O(\tilde{c}^2) \\
= -F_{ij}(\theta) + O(c^2) + O(\tilde{c}^2). \tag{2.12}
\]

The last equality follows from the fact that the FIM, $F_n(\theta)$, is symmetric.
The variance of \( \hat{H}^{(L)}_{ij} \) is to be computed next. It is given by,

\[
\text{var}[\hat{H}^{(L)}_{ij}|\theta] = \frac{1}{4} \text{var}[(\hat{J}^{(L)}_{ij} + \hat{J}^{(L)}_{ji})|\theta] \\
= \frac{1}{4} \left( \text{var}[\hat{J}^{(L)}_{ij}|\theta] + \text{var}[\hat{J}^{(L)}_{ji}|\theta] + 2\text{cov}[\hat{J}^{(L)}_{ij}, \hat{J}^{(L)}_{ji}|\theta] \right). \tag{2.13}
\]

The expression of a typical variance term, \( \text{var}[\hat{J}^{(L)}_{ij}|\theta] \), in (2.13) would now be determined followed by the deduction of the expression of covariance term, \( \text{cov}[\hat{J}^{(L)}_{ij}, \hat{J}^{(L)}_{ji}|\theta] \).

By the use of (2.10), \( \text{var}[\hat{J}^{(L)}_{ij}|\theta] \) is given by,

\[
\text{var}[\hat{J}^{(L)}_{ij}|\theta] = \text{var}\left[ \left( \sum_{l,m} H_{lm}(\theta|Z) \frac{\Delta_m}{\Delta_j} \frac{\Delta_i}{\Delta_j} + O_{\Delta\Delta Z}(c^2) + O_{\Delta\Delta Z}(\tilde{c}) \right) \right] \\
= \text{var}\left[ \sum_{l,m} H_{lm}(\theta|Z) \frac{\Delta_m}{\Delta_j} \frac{\Delta_i}{\Delta_j} \right] + \text{var} [O_{\Delta\Delta Z}(c^2)] + \text{var} [O_{\Delta\Delta Z}(\tilde{c})] \\
+ \text{var} [O_{\Delta\Delta Z}(\tilde{c}^2)] + 2\text{cov} [O_{\Delta\Delta Z}(1), O_{\Delta\Delta Z}(c^2)] \\
+ 2\text{cov} [O_{\Delta\Delta Z}(1), O_{\Delta\Delta Z}(\tilde{c})] + 2\text{cov} [O_{\Delta\Delta Z}(1), O_{\Delta\Delta Z}(\tilde{c}^2)] \\
+ 2\text{cov} [O_{\Delta\Delta Z}(c^2), O_{\Delta\Delta Z}(\tilde{c})] + 2\text{cov} [O_{\Delta\Delta Z}(c^2), O_{\Delta\Delta Z}(\tilde{c}^2)] \\
+ 2\text{cov} [O_{\Delta\Delta Z}(\tilde{c}), O_{\Delta\Delta Z}(\tilde{c}^2)], \tag{2.14}
\]
in which \( O_{\Delta\Delta Z}(1) \) represents the term \( \sum_{l,m} H_{lm}(\theta|Z)(\Delta_m/\Delta_j)(\Delta_i/\Delta_j) \). Note that the variance terms, \( \text{var} [O_{\Delta\Delta Z}(c^2)] \), \( \text{var} [O_{\Delta\Delta Z}(\tilde{c})] \) and \( \text{var} [O_{\Delta\Delta Z}(\tilde{c}^2)] \), respectively, turn out to be \( O(c^4) \), \( O(\tilde{c}^2) \) and \( O(\tilde{c}^4) \) by C.4. While the covariance terms, \\
\[
2\text{cov} [O_{\Delta\Delta Z}(1), O_{\Delta\Delta Z}(c^2)], 2\text{cov} [O_{\Delta\Delta Z}(1), O_{\Delta\Delta Z}(\tilde{c})], 2\text{cov} [O_{\Delta\Delta Z}(1), O_{\Delta\Delta Z}(\tilde{c}^2)] \\
and 2\text{cov} [O_{\Delta\Delta Z}(c^2), O_{\Delta\Delta Z}(\tilde{c})], 2\text{cov} [O_{\Delta\Delta Z}(c^2), O_{\Delta\Delta Z}(\tilde{c}^2)],
\]
respectively, turn out to be \( O(c^2) \), \( O(\tilde{c}^2) \) and \( O(\tilde{c}^2 \tilde{c}^2) \) by C.4, a few other covariance terms — \\
\[
2\text{cov} [O_{\Delta\Delta Z}(1), O_{\Delta\Delta Z}(\tilde{c})], 2\text{cov} [O_{\Delta\Delta Z}(c^2), O_{\Delta\Delta Z}(\tilde{c})],
\]
and \( 2\text{cov} [O_{\Delta\Delta Z}(c^2), O_{\Delta\Delta Z}(\tilde{c})] \), respectively.
\(O_{\Delta, \Delta Z}(c) | \theta\) — vanish since, for any given \(i\), they involve terms like \(E[(\tilde{\Delta}_l \tilde{\Delta}_{m2} \tilde{\Delta}_{l2}) / \tilde{\Delta}_j^2]\), \(l_1, l_2, m_2 = 1, \cdots, p\), all of which are zero by implication I and rest of the terms appeared in \(O_{\Delta, \Delta Z}(c)\) and \(O_{\Delta, \Delta Z}(c^2)\) do not depend on \(\Delta\) (as already mentioned below (2.8) that \(\theta_1^\pm\) in \(O_{\Delta, \Delta Z}(c^2)\) and \(\theta_2^\pm\) in \(O_{\Delta, \Delta Z}(c)\) depend only on \(\theta\) and \(\theta + c\Delta\)). Although the remaining term, 2cov \([O_{\Delta, \Delta Z}(c), O_{\Delta, \Delta Z}(c^2)]\theta\), in (2.14) involves terms like \(E[(\tilde{\Delta}_{m1} \tilde{\Delta}_{l1} \tilde{\Delta}_{m2} \tilde{\Delta}_{l2}) / \tilde{\Delta}_j^2]\), \(l_1, m_1, l_2, m_2 = 1, \cdots, p\), which are zero by implication I, the ‘big-O’ term, \(O_{\Delta, \Delta Z}(c^2)\), contains \(\overline{\theta}\) that depends on \(\theta\) and \(\theta + c\Delta\) (see below (2.6)). Therefore, 2cov \([O_{\Delta, \Delta Z}(c), O_{\Delta, \Delta Z}(c^2)]\theta\] is not zero and turns out to be \(O(c^3)\) by C.4. Then, by absorbing the term \(O(c^4)\) into \(O(c^2)\) and both \(O(c^4)\) and \(O(c^3)\) into \(O(c^2)\), the variance, \(\text{var}[\hat{J}^{(l)}_{ij}]\), turns out to be,

\[
\begin{align*}
\text{var}[\hat{J}^{(l)}_{ij} | \theta] &= \text{var} \left[ \sum_{l,m} H_{lm}(\theta|z) \frac{\Delta_m}{\Delta_j} \frac{\Delta_l}{\Delta_i} \bigg| \theta \right] + O(c^2) + O(c^2) + O(c^2c^2) \\
&= \sum_{l_1, m_1, l_2, m_2} E \left[ \{H_{l_1 m_1}(\theta|z) H_{l_2 m_2}(\theta|z)\} | \theta \right] E \left[ \frac{\Delta_{m_1} \Delta_{m_2}}{\Delta_j^2} \right] E \left[ \frac{\Delta_{l_1} \Delta_{l_2}}{\Delta_i^2} \right] \\
&\quad - (E [H_{ij}(\theta|z)] \bigg| \theta \bigg)^2 + O(c^2) + O(c^2) + O(c^2c^2) \\
&= \sum_{l,m} E \left[ \{H_{lm}(\theta|z)^2 \bigg| \theta \right] E \left[ \frac{\Delta_{m_1}^2}{\Delta_j^2} \right] E \left[ \frac{\Delta_{l_2}^2}{\Delta_i^2} \right] - (E [H_{ij}(\theta|z)] \bigg| \theta \bigg)^2 \\
&\quad + O(c^2) + O(c^2) + O(c^2c^2) \\
&= \sum_{l,m} \{ \text{var} [H_{lm}(\theta|z)] | \theta \} + (E [H_{lm}(\theta|z)] \bigg| \theta \bigg)^2 \right] E \left[ \frac{\Delta_{m_1}^2}{\Delta_j^2} \right] E \left[ \frac{\Delta_{l_2}^2}{\Delta_i^2} \right] \\
&\quad - (E [H_{ij}(\theta|z)] \bigg| \theta \bigg)^2 + O(c^2) + O(c^2) + O(c^2c^2)
\end{align*}
\]
\[ = \sum_{l,m} a_{lm}(i,j) \text{var} \left[ H_{lm}(\theta|Z) \right] \theta + \sum_{l,m} a_{lm}(i,j) \left( E \left[ H_{lm}(\theta|Z) \right] \right)^2 \]
\[ + O(c^2) + O(\bar{c}^2) + O(c^2\bar{c}^2) \quad (2.15) \]
in which \( a_{lm}(i,j) = E[\Delta_m^2/\Delta_j^2]E[\Delta_l^2/\Delta_i^2] \). The first term after the third equality follows from the fact that \( \Delta, \bar{\Delta}, \Delta \) and \( Z \) are statistically independent of each other and the second term has already been seen during the derivation of \( E[\bar{J}^{(L)}_{ij}|\theta] \) in (2.11). The fourth equality follows by condition C.1 and implication I. Note that \( \text{var}[\hat{H}^{(L)}_{ii}|\theta] \) follows straight from the expression of \( \text{var}[\hat{J}^{(L)}_{ij}|\theta] \) by replacing \( j \) with \( i \) in (2.15).

Next, the expression of \( \text{cov}[\hat{J}^{(L)}_{ij}, \hat{J}^{(L)}_{ji}|\theta], j \neq i \), would be deduced. By using an identical argument to the one that is mentioned right after (2.14), it can be readily shown that \( \text{cov}[\hat{J}^{(L)}_{ij}, \hat{J}^{(L)}_{ji}|\theta] \) is given by,
\[
\text{cov}[\hat{J}^{(L)}_{ij}, \hat{J}^{(L)}_{ji}|\theta] = \text{cov} \left[ \left( \sum_{l,m} H_{lm}(\theta|Z) \frac{\Delta_m}{\Delta_j} \frac{\Delta_l}{\Delta_i} \right), \left( \sum_{l,m} H_{lm}(\theta|Z) \frac{\Delta_m}{\Delta_i} \frac{\Delta_l}{\Delta_j} \right) \right] \theta \]
\[ + O(c^2) + O(\bar{c}^2) + O(c^2\bar{c}^2). \quad (2.16) \]

Further simplification of the first term in the right-hand-side yields,
\[
\text{cov} \left[ \left( \sum_{l,m} H_{lm}(\theta|Z) \frac{\Delta_m}{\Delta_j} \frac{\Delta_l}{\Delta_i} \right), \left( \sum_{l,m} H_{lm}(\theta|Z) \frac{\Delta_m}{\Delta_i} \frac{\Delta_l}{\Delta_j} \right) \right] \theta
\]
\[ = E \left[ \left( \sum_{l,m} H_{lm}(\theta|Z) \frac{\Delta_m}{\Delta_j} \frac{\Delta_l}{\Delta_i} \right) \left( \sum_{l,m} H_{lm}(\theta|Z) \frac{\Delta_m}{\Delta_i} \frac{\Delta_l}{\Delta_j} \right) \right] \theta
\[ \quad - E \left[ H_{ij}(\theta|Z) \right] \theta E \left[ H_{ji}(\theta|Z) \right] \theta
\]
\[ = \sum_{l_1,m_1,l_2,m_2} E \left[ \{ H_{l_1m_1}(\theta|Z)H_{l_2m_2}(\theta|Z) \} \right] \theta \left[ E \left[ \frac{\Delta_{m_1}}{\Delta_i} \frac{\Delta_{m_2}}{\Delta_j} \right] \right] \theta
\[ \quad - \left( F_{ij}(\theta) \right)^2
\]
\[ = \left\{ E \left[ H_{ij}(\theta|Z)H_{ji}(\theta|Z) \right] \theta + E \left[ H_{ii}(\theta|Z)H_{jj}(\theta|Z) \right] \theta + E \left[ H_{jj}(\theta|Z)H_{ii}(\theta|Z) \right] \theta
\[ + E \left[ H_{ji}(\theta|Z)H_{ij}(\theta|Z) \right] \theta \right\} \theta
\[ \quad - \left( F_{ij}(\theta) \right)^2
\]
28
Here, the first term after the second equality follows from the fact that \( \tilde{\Delta}, \Delta \) and \( Z \) are statistically independent of each other and the second term after the second equality follows by the Hessian-based definition and symmetry of FIM. The third equality follows by the use of the fact that \( j \neq i \) and by condition C.1 and implication I.

Therefore, \( \text{cov}[\hat{J}^{(L)}_{ij}, \hat{J}^{(L)}_{ji}|\theta] \) turns out to be,

\[
\text{cov}[\hat{J}^{(L)}_{ij}, \hat{J}^{(L)}_{ji}|\theta] = 2 \{ \text{var}[\hat{H}^{(L)}_{ij}|\theta] + (E[\hat{H}^{(L)}_{ij}|\theta]^2) + 2E[H_{ii}(\theta)H_{jj}(\theta)|\theta] - (F_{ij}(\theta))^2 \}.
\]  

(2.18)

Now, the variance of \( \hat{H}^{(L)}_{ij} \), \( \text{var}[\hat{H}^{(L)}_{ij}|\theta] \), for \( j \neq i \), can be readily obtained from (2.13) by using (2.15) and (2.18). As already indicated, \( \text{var}[\hat{J}^{(L)}_{ii}|\theta] \) is same as \( \text{var}[\hat{J}^{(L)}_{ii}|\theta] \) that can be directly obtained from (2.15) by replacing \( j \) with \( i \). Since the primary objective of the current work is to show that the variance of the new estimators, \( \hat{F}_{ij} \), of the elements, \( F_{ij}(\theta) \), of \( F_n(\theta) \) are less than that of the current estimators, \( \hat{F}_{ij} \), it would suffice to show that \( \text{var}[\hat{H}^{(L)}_{ij}|\theta] \) is less than \( \text{var}[\hat{H}^{(L)}_{ij}|\theta] \). This later fact can be proved by comparing the contributions of the variance and covariance terms to \( \text{var}[\tilde{H}^{(L)}_{ij}|\theta] \) with the contributions of the respective variance and
covariance terms (as appeared in (2.13)) to \( \var[H_{ij}] \). This is the task of the next section.

Mean and Variance of \( \tilde{H}_{ij}^{(L)} \)

We now derive the mean and variance of the element of \( \tilde{H}_{ij}^{(L)} \) for the purpose of contrasting with the same quantities for \( \hat{H}_{ij}^{(L)} \). Consider the \((i, j)\)-th element of \( \tilde{H}_k \) associated with (2.3) that is given by,

\[
\tilde{H}_{ij}^{(L)} = \hat{H}_{ij}^{(L)} - \frac{1}{2} \sum_l \sum_{m \in I_l} \left[ (-F_{lm}(\theta)) \frac{\Delta m}{\Delta_j} \frac{\tilde{\Delta}_l}{\Delta_i} + (-F_{lm}(\theta)) \frac{\Delta m}{\Delta_j} \frac{\tilde{\Delta}_l}{\Delta_i} \right], \quad \forall j \in I_i^c
\]

and

\[
\tilde{H}_{ij}^{(L)} = -F_{ij}(\theta), \quad \forall j \in I_i. \tag{2.20}
\]

In (2.19), \( \hat{J}_{ij}^{(L)} \) is defined as,

\[
\hat{J}_{ij}^{(L)} = \hat{J}_{ij}^{(L)} - \sum_l \sum_{m \in I_l} (-F_{lm}(\theta)) \frac{\Delta m}{\Delta_j} \frac{\tilde{\Delta}_l}{\Delta_i}, \quad \forall j \in I_i^c. \tag{2.21}
\]

Note that \( \forall j \in I_i^c \) in (2.21),

\[
E \left[ \sum_l \sum_{m \in I_l} (-F_{lm}(\theta)) \frac{\Delta m}{\Delta_j} \frac{\tilde{\Delta}_l}{\Delta_i} \bigg| \theta \right] = 0,
\]

implying that \( E[\tilde{J}_{ij}^{(L)}] = E[\hat{J}_{ij}^{(L)}], \forall j \in I_i^c \). By using this fact along with (2.19), (2.11) and the symmetry of the FIM (for all \( j \in I_i^c \)) and by using (2.20) (for all \( j \in I_i \)), the
following can be obtained readily, \( \forall i = 1, \cdots, p, \)
\[
E[\tilde{H}^{(L)}_{ij}|\theta] = \begin{cases} 
- F_{ij}(\theta) + O(c^2) + O(\tilde{c}^2), & \forall j \in I^c_i, \\
- F_{ij}(\theta), & \forall j \in I_i. 
\end{cases} \tag{2.22}
\]

Noticeably, the expressions of the ‘big-O’ terms both in (2.12) and in the first equation of (2.22) are precisely same implying that \( E[\tilde{H}^{(L)}_{ij}|\theta] = E[\hat{H}^{(L)}_{ij}|\theta], \forall j \in I^c_i. \)

While \( \text{var}[\tilde{H}^{(L)}_{ij}|\theta] = 0, \forall j \in I_i, \) clearly implying that \( \text{var}[\tilde{H}^{(L)}_{ij}|\theta] < \text{var}[\hat{H}^{(L)}_{ij}|\theta], \forall j \in I^c_i, \) is the task that will be considered now. In fact, this is the main result associated with the variance reduction from prior information available in terms of the known elements of \( F_n(\theta). \)

The first step in determining this variance is to note that the expression of \( \hat{J}^{(L)}_{ij} \) in (2.10) can be decomposed into two parts as shown below,
\[
\hat{J}^{(L)}_{ij} = \sum_l \left[ \sum_{m \in I^c_i} H_{lm}(\theta|Z) \frac{\Delta m}{\Delta_j} \frac{\Delta_l}{\Delta_i} + \sum_{m \in I_i} H_{lm}(\theta|Z) \frac{\Delta m}{\Delta_j} \frac{\Delta_l}{\Delta_i} \right] + O_{\Delta,\Delta Z}(c^2) + O_{\Delta,\Delta Z}(\tilde{c}) + O_{\Delta,\Delta Z}(\tilde{c}^2). \tag{2.23}
\]

The elements, \( H_{lm}(\theta|Z), \) of \( H(\theta|Z) \) in the right-hand-side of (2.23) are not known. However, since by (Hessian-based) definition \( E[H_{lm}(\theta|Z)|\theta] = -F_{lm}(\theta), \) approximation of the unknown elements of \( H(\theta|Z) \) in the right-hand-side of (2.23), particularly those elements that correspond to the elements of the FIM that are known \textit{a priori}, by the negative of those elements of \( F_n(\theta) \) is the primary idea based on which the modified resampling algorithm is developed. This approximation introduces an error term, \( e_{lm}(\theta|Z), \) that can be defined by,
\[
H_{lm}(\theta|Z) = -F_{lm}(\theta) + e_{lm}(\theta|Z), \quad \forall m \in I_i, \ l = 1, \cdots, p, \tag{2.24}
\]

31
and this error term satisfies the following two conditions that directly follow from (2.24), \( \forall m \in \mathbb{I}_l, l = 1, \ldots, p, \)

\[
E[e_{lm}(\theta|Z)|\theta] = 0, \tag{2.25}
\]

\[
\text{var}[e_{lm}(\theta|Z)|\theta] = \text{var}[H_{lm}(\theta|Z)|\theta]. \tag{2.26}
\]

Substitution of (2.24) in (2.23) results in a known part in the right-hand-side of (2.23) involving the analytically known elements of FIM. This known part is transferred to the left-hand-side of (2.23) and, consequently, acts as a feedback to the current resampling algorithm yielding, in the process, \( \tilde{J}_{ij}^{(L)} \),

\[
\tilde{J}_{ij}^{(L)} = \sum_{l,m} X_{lm}(\theta|Z) \frac{\Delta m}{\Delta_j} \frac{\Delta l}{\Delta_i} + O_{\Delta_{ij}Z}(c^2) + O_{\Delta_{ij}Z}(\tilde{c}) + O_{\Delta_{ij}Z}(\tilde{e}^2), \quad \forall j \in \mathbb{I}_i^c. \tag{2.27}
\]

This expression of \( \tilde{J}_{ij}^{(L)} \) can be employed to show that the new estimator, \( \tilde{H}_{ij}^{(L)} \), in (2.19) has less variance than that of the current estimator, \( \hat{H}_{ij}^{(L)} \). This is achieved in the following.

Introduce \( X_{lm}, l = 1, \ldots, p \), as defined below,

\[
X_{lm}(\theta|Z) = \begin{cases} 
  e_{lm}(\theta|Z), & \text{if } m \in \mathbb{I}_l, \\
  H_{lm}(\theta|Z), & \text{if } m \in \mathbb{I}_l^c.
\end{cases} \tag{2.28}
\]

Then, (2.27) can be compactly written as,

\[
\tilde{J}_{ij}^{(L)} = \sum_{l,m} X_{lm}(\theta|Z) \frac{\Delta m}{\Delta_j} \frac{\Delta l}{\Delta_i} + O_{\Delta_{ij}Z}(c^2) + O_{\Delta_{ij}Z}(\tilde{c}) + O_{\Delta_{ij}Z}(\tilde{e}^2), \quad \forall j \in \mathbb{I}_i^c. \tag{2.29}
\]
On the other hand, variance of $\tilde{H}_{ij}^{(L)}$ in (2.19) is given by,

$$\text{var}[\tilde{H}_{ij}^{(L)}|\theta] = \frac{1}{4} \text{var}[(\tilde{J}_{ij}^{(L)} + \tilde{J}_{ji}^{(L)})|\theta]$$

$$= \frac{1}{4} \left( \text{var}[\tilde{J}_{ij}^{(L)}|\theta] + \text{var}[\tilde{J}_{ji}^{(L)}|\theta] + 2\text{cov}[\tilde{J}_{ij}^{(L)}, \tilde{J}_{ji}^{(L)}|\theta] \right).$$

(2.30)

Therefore, $\text{var}[\tilde{J}_{ij}^{(L)}|\theta]$ and $\text{cov}[\tilde{J}_{ij}^{(L)}, \tilde{J}_{ji}^{(L)}|\theta]$ are required to be computed next in order to compare them, respectively, with $\text{var}[\hat{J}_{ij}^{(L)}|\theta]$ and $\text{cov}[\hat{J}_{ij}^{(L)}, \hat{J}_{ji}^{(L)}|\theta]$ in (2.18) and, consequently, facilitating the comparison between $\text{var}[\tilde{H}_{ij}^{(L)}|\theta]$ and $\text{var}[\hat{H}_{ij}^{(L)}|\theta]$, $\forall j \in \mathbb{I}_i^c$.

The variance of $\tilde{J}_{ij}^{(L)}$, $\forall j \in \mathbb{I}_i^c$, can be computed by considering the right-hand-side of (2.29) in an identical way as described for $\tilde{J}_{ij}^{(L)}$ in the previous subsection. The expression for $\text{var}[\tilde{J}_{ij}^{(L)}|\theta]$, $\forall j \in \mathbb{I}_i^c$, follows readily from (2.15) by replacing $H_{lm}$ with $X_{lm}$ because of the similarity between (2.10) and (2.29) as shown below, $\forall j \in \mathbb{I}_i^c$, $i = 1, \cdots, p$,

$$\text{var}[\tilde{J}_{ij}^{(L)}|\theta] = \sum_{l,m} a_{lm}(i, j) \text{var} [X_{lm}(\theta|Z)|\theta] + \sum_{l,m \neq ij} a_{lm}(i, j) (E [X_{lm}(\theta|Z)|\theta])^2$$

$$+ O(c^2) + O(c^2) + O(c^2)$$

$$= \sum_{l,m} a_{lm}(i, j) \text{var} [H_{lm}(\theta|Z)|\theta] + \sum_{l,m \neq ij} a_{lm}(i, j) (E [H_{lm}(\theta|Z)|\theta])^2$$

$$+ O(c^2) + O(c^2) + O(c^2).$$

(2.31)

The second equality follows by the use of (2.28) and a subsequent use of (2.25)-(2.26) on the resulting expression.
Subtracting (2.31) from (2.15), it can be readily seen that $\forall j \in \mathbb{I}_i^c$,

$$\text{var}[\tilde{J}_{ij}^{(L)}|\theta] - \text{var}[\hat{J}_{ij}^{(L)}|\theta] = \sum_l \sum_{m \in \mathbb{I}_i} a_{lm}(i, j) (E[H_{lm}(\theta|\mathbf{Z})|\theta])^2$$

$$+ O(c^2) + O(\bar{c}^2) + O(c^2 \bar{c}^2)$$

$$= \sum_l \sum_{m \in \mathbb{I}_i} a_{lm}(i, j) (F_{lm}(\theta))^2$$

$$+ O(c^2) + O(\bar{c}^2) + O(c^2 \bar{c}^2) > 0. \quad (2.32)$$

Here, the second equality follows by the Hessian-based definition of FIM, $E[H_{lm}(\theta|\mathbf{Z})|\theta] = -F_{lm}(\theta)$, and the inequality follows from the fact that $a_{lm}(i, j) = (E[\Delta_m^2/\Delta_j^2] E[\bar{\Delta}_l^2/\bar{\Delta}_i^2]) > 0$, $l, m = 1, \cdots, p$, for any given $(i, j)$ and assuming that at least one of the known elements, $F_{lm}(\theta)$, in (2.32) is not equal to zero. It must be remarked that the bias terms, $O(c^2)$, $O(\bar{c}^2)$ and $O(c^2 \bar{c}^2)$, can be made negligibly small by selecting $c$ and $\bar{c}$ small enough that are primarily controlled by users. Note that if $\Delta_1, \cdots, \Delta_p$ and $\bar{\Delta}_1, \cdots, \bar{\Delta}_p$ are both assumed to be Bernoulli ±1 i.i.d. random variables, then $a_{lm}(i, j)$ turns out to be unity.

At this point it is already clear that $\text{var}[\bar{H}_{ii}^{(L)}|\theta] < \text{var}[\hat{H}_{ii}^{(L)}|\theta]$ if $j = i \in \mathbb{I}_i^c$, since $\text{var}[\bar{H}_{ii}^{(L)}|\theta]$ and $\text{var}[\hat{H}_{ii}^{(L)}|\theta]$ are exactly given by $\text{var}[\tilde{J}_{ii}^{(L)}|\theta]$ and $\text{var}[\hat{J}_{ii}^{(L)}|\theta]$, respectively, and $\text{var}[\tilde{J}_{ii}^{(L)}|\theta] < \text{var}[\hat{J}_{ii}^{(L)}|\theta]$ by (2.32).

Next step is to compare $\text{cov}[\tilde{J}_{ij}^{(L)}, \tilde{J}_{ji}^{(L)}|\theta]$ to $\text{cov}[\hat{J}_{ij}^{(L)}, \hat{J}_{ji}^{(L)}|\theta]$, $j \neq i$, $\forall j \in \mathbb{I}_i^c$, in order to conclude that $\text{var}[\bar{H}_{ij}^{(L)}|\theta] < \text{var}[\hat{H}_{ij}^{(L)}|\theta]$. As $\text{var}[\tilde{J}_{ij}^{(L)}|\theta]$ is deduced from $\text{var}[\tilde{J}_{ij}^{(L)}|\theta]$ by the similarity of the expressions between (2.10) and (2.29), the expression of $\text{cov}[\tilde{J}_{ij}^{(L)}, \tilde{J}_{ji}^{(L)}|\theta]$ also follows readily by replacing $H_{lm}$ with $X_{lm}$ and using
identical arguments that are used in deducing (2.16) and (2.17). The expression of $\text{cov}[\tilde{J}^{(L)}_{ij}, \tilde{J}^{(L)}_{ji}|\theta]$, therefore, turns out to be, for $j \neq i, \forall j \in I^c, i = 1, \ldots, p,$

$$\text{cov}[\tilde{J}^{(L)}_{ij}, \tilde{J}^{(L)}_{ji}|\theta] = 2 \left\{ \var \left[ (X_{ij}(\theta|Z)|\theta) + (E[(X_{ij}(\theta|Z)|\theta)]^2 \right] 
+ 2E[X_{ii}(\theta|Z)X_{jj}(\theta|Z)|\theta] - (F_{ij}(\theta))^2 + O(c^2) + O(\tilde{c}^2) + O(c^2\tilde{c}^2)
= 2 \left\{ \var \left[ (H_{ij}(\theta|Z)|\theta) + (E[(H_{ij}(\theta|Z)|\theta)]^2 \right] 
+ 2E[X_{ii}(\theta|Z)X_{jj}(\theta|Z)|\theta] - (F_{ij}(\theta))^2 + O(c^2) + O(\tilde{c}^2) + O(c^2\tilde{c}^2). \right.$$  

(2.33)

Since $j \in I^c$, the second equality follows from the definition of $X_{ij}$ in (2.28). In (2.33), $X_{ii}$ and $X_{jj}$ must take the form from one of following four possibilities: (1) $e_{ii}$ and $e_{jj}$, (2) $e_{ii}$ and $H_{jj}$, (3) $H_{ii}$ and $e_{jj}$, (4) $H_{ii}$ and $H_{jj}$. Subtracting (2.33) from (2.18), the difference between $\text{cov}[\tilde{J}^{(L)}_{ij}, \tilde{J}^{(L)}_{ji}|\theta]$ and $\text{cov}[\tilde{J}^{(L)}_{ij}, \tilde{J}^{(L)}_{ji}|\theta]$ is given by,

$$\text{cov}[\tilde{J}^{(L)}_{ij}, \tilde{J}^{(L)}_{ji}|\theta] - \text{cov}[\tilde{J}^{(L)}_{ij}, \tilde{J}^{(L)}_{ji}|\theta] = 2 \left( E[H_{ii}(\theta|Z)H_{jj}(\theta|Z)|\theta] - E[X_{ii}(\theta|Z)X_{jj}(\theta|Z)|\theta] 
+ O(c^2) + O(\tilde{c}^2) + O(c^2\tilde{c}^2), \quad \text{for } j \neq i, \forall j \in I^c. \right.$$

(2.34)

For the first possibility, it is shown below that $E[H_{ii}(\theta|Z)H_{jj}(\theta|Z)|\theta] - E[X_{ii}(\theta|Z)X_{jj}(\theta|Z)|\theta] X_{jj}(\theta|Z)|\theta]$ is given by $F_{ii}(\theta)F_{jj}(\theta)$ that is greater than 0 since $F_n(\theta)$ is positive definite matrix and a positive definite matrix always has non-zero positive diagonal entries.

$$E[H_{ii}(\theta|Z)H_{jj}(\theta|Z)|\theta] - E[X_{ii}(\theta|Z)X_{jj}(\theta|Z)|\theta]$$

35
\[
\begin{align*}
\quad & = E \left[ H_{ii}(\theta|Z) H_{jj}(\theta|Z) | \theta \right] - E \left[ e_{ii}(\theta|Z) e_{jj}(\theta|Z) | \theta \right] \\
\quad & = E \left[ H_{ii}(\theta|Z) H_{jj}(\theta|Z) | \theta \right] - E \left[ (H_{ii}(\theta|Z) + F_{ii}(\theta))(H_{jj}(\theta|Z) + F_{jj}(\theta)) | \theta \right] \\
\quad & = F_{ii}(\theta) F_{jj}(\theta) \\
\quad & > 0.
\end{align*}
\]

Here, the second equality follows by (2.24) and the third equality follows by the Hessian-based definition of FIM, \( E[H_{lm}(\theta|Z)|\theta] = -F_{lm}(\theta) \). While the above difference is 0 for the fourth possibility (\( X_{ii} \) and \( X_{jj} \), respectively, being same as \( H_{ii} \) and \( H_{jj} \)), it can be shown easily in a similar way as determined above for the first possibility that the difference again becomes \( F_{ii}(\theta) F_{jj}(\theta) > 0 \) for the second (\( X_{ii} \) and \( X_{jj} \), respectively, being same \( e_{ii} \) and \( H_{jj} \)) and third (\( X_{ii} \) and \( X_{jj} \), respectively, being same as \( H_{ii} \) and \( e_{jj} \)) possibilities.

Therefore, using (2.32) and the fact that \( E[H_{ii}(\theta|Z) H_{jj}(\theta|Z)|\theta] - E[X_{ii}(\theta|Z) X_{jj}(\theta|Z)|\theta] \geq 0 \), that appeared in (2.34), and also noting that the bias terms, \( O(c^2) \), \( O(\tilde{c}^2) \) and \( O(c^2\tilde{c}^2) \), can be made negligibly small by selecting \( c \) and \( \tilde{c} \) small enough, the following can be concluded immediately from (2.13) and (2.30),

\[
\text{var}[\tilde{H}^{(L)}_{ij}|\theta] < \text{var}[\hat{H}^{(L)}_{ij}|\theta], \quad i, j = 1, \ldots, p.
\]

Here, the ranges of \( i \) and \( j \) are mentioned as \( \{1, \ldots, p\} \) not just \( \mathbb{I}_j \) since for the other case, \( \mathbb{I}_i \), it has already been noted that \( \text{var}[\tilde{H}^{(L)}_{ij}|\theta] = 0 \), by (2.20), that is less than \( \text{var}[\hat{H}^{(L)}_{ij}|\theta] \).

The objective of the next section is to show that the variance of \( \tilde{H}^{(g)}_{ij} \) is less than
the variance of $\hat{H}_{ij}^{(g)}$.

### 2.3.2 Case 2: measurements of $g$ are available

The mean and variance of $\hat{H}_{ij}^{(g)}$ are computed first followed by the mean and variance of $\tilde{H}_{ij}^{(g)}$.

Mean and Variance of $\hat{H}_{ij}^{(g)}$

In this section, it is assumed that the measurements of exact gradient vector, $g$, are available. Therefore, the $(i, j)$-th element, $\hat{H}_{ij}^{(g)}$, (the dependence on $k$ is suppressed) of $\hat{H}_{k}^{(g)}$ is directly given by,

$$\hat{H}_{ij}^{(g)} = \frac{1}{2} \left[ \frac{g_i(\theta + c\Delta|Z) - g_i(\theta - c\Delta|Z)}{2c\Delta_j} + \frac{g_j(\theta + c\Delta|Z) - g_j(\theta - c\Delta|Z)}{2c\Delta_i} \right]. \tag{2.35}$$

As earlier, consider a typical term (say, the first term in the right-hand-side of (2.35)) and denote it by $\hat{j}_{ij}^{(g)}$ which, based on a Taylor expansion, turns out to be as follows,

$$\hat{j}_{ij}^{(g)} = g_i(\theta + c\Delta|Z) - g_i(\theta - c\Delta|Z)$$

$$= \frac{1}{2c\Delta_j} \left[ \left( g_i(\theta|Z) + c \sum_l H_{il}(\theta|Z) \Delta_l + \frac{1}{2} c^2 \sum_{l,m} \frac{\partial H_{il}(\theta|Z)}{\partial \theta_m} \Delta_m \Delta_l \right) - \left( g_i(\theta|Z) + \left( -c \right) \sum_l H_{il}(\theta|Z) \Delta_l \right) \right. \right.$$

$$\left. \left. + \frac{1}{6} c^3 \sum_{l,m,s} \frac{\partial^2 H_{il}(\theta^1|Z)}{\partial \theta_s \partial \theta_m} \Delta_s \Delta_m \Delta_l \right) - \left( g_i(\theta|Z) + \left( -c \right) \sum_l H_{il}(\theta|Z) \Delta_l \right) \right. \right.$$

$$\left. \left. + \frac{1}{2} \left( -c \right)^2 \sum_{l,m} \frac{\partial H_{il}(\theta|Z)}{\partial \theta_m} \Delta_m \Delta_l + \frac{1}{6} \left( -c \right)^3 \sum_{l,m,s} \frac{\partial^2 H_{il}(\theta^1|Z)}{\partial \theta_s \partial \theta_m} \Delta_s \Delta_m \Delta_l \right) \right]$$

$$= \sum_l H_{il}(\theta|Z) \Delta_l \Delta_j + \frac{1}{12} c^2 \sum_{l,m,s} \left( \frac{\partial^2 H_{il}(\theta^1|Z)}{\partial \theta_s \partial \theta_m} + \frac{\partial^2 H_{il}(\theta^1|Z)}{\partial \theta_s \partial \theta_m} \right) \Delta_s \Delta_m \Delta_l \Delta_j. \tag{2.36}$$
Here, the point of evaluation, $\theta$, is suppressed in the random \textit{‘big-O’} term, $O_{\Delta Z}(c^2)$ as done in previous section, $\theta^\pm_1$ denote some points on the line segments between $\theta$ and $\theta \pm c\Delta$ and, the superscript, $\pm$ sign, in $\theta^\pm_1$, corresponds to whether $\theta + c\Delta$ or $\theta - c\Delta$ is being considered for the argument of $g_i(\cdot)$ above. The expectation of $\hat{J}_{ij}^{(g)}$ follows straight from (2.36) by using condition C.1,

$$E[\hat{J}_{ij}^{(g)}|\theta] = E[H_{ij}(\theta|Z)|\theta] + O(c^2),$$

yielding the expectation of $\hat{H}_{ij}^{(g)}$ as,

$$E[\hat{H}_{ij}^{(g)}|\theta] = \frac{1}{2} \left( E[\hat{J}_{ij}^{(g)}|\theta] + E[\hat{J}_{ji}^{(g)}|\theta] \right) = -F_{ij}(\theta) + O(c^2),$$

in which the Hessian-based definition, $E[H_{ij}(\theta|Z)|\theta] = -F_{ij}(\theta)$, and the symmetry, $F_{ij}(\theta) = F_{ji}(\theta)$, of FIM are used.

Next, the variance of $\hat{H}_{ij}^{(g)}$ is considered and it is given by,

$$\text{var}[\hat{H}_{ij}^{(g)}|\theta] = \frac{1}{4} \text{var}[\hat{J}_{ij}^{(g)} + \hat{J}_{ji}^{(g)}]|\theta]$$

$$= \frac{1}{4} \left( \text{var}[\hat{J}_{ij}^{(g)}|\theta] + \text{var}[\hat{J}_{ji}^{(g)}|\theta] + 2\text{cov}[\hat{J}_{ij}^{(g)}, \hat{J}_{ji}^{(g)}|\theta] \right).$$

The expression of the variance term and the covariance term in (2.39) will now be determined in succession.

By the use of (2.36), $\text{var}[\hat{J}_{ij}^{(g)}|\theta]$ in (2.39) is given by,

$$\text{var}[\hat{J}_{ij}^{(g)}|\theta] = \text{var} \left[ \left( \sum_l H_{il}(\theta|Z) \frac{\Delta_l}{\Delta_j} + O_{\Delta Z}(c^2) \right) \theta \right]$$

$$= \text{var} \left[ \sum_l H_{il}(\theta|Z) \frac{\Delta_l}{\Delta_j} \theta \right] + O(c^2),$$

38
in which \( \text{var}[O_{\Delta, Z}(c^2)|\theta] \), which is a \( O(c^4) \) term by C.4, is absorbed in the \( 2\text{cov}[\sum_l H_{il}(\theta|Z)(\Delta_l/\Delta_j), O_{\Delta, Z}(c^2)|\theta] \) that is a \( O(c^2) \) term (by C.4). Further, \( \text{var}[\hat{J}_{ij}^{(g)}|\theta] \) can be written as,

\[
\text{var}[\hat{J}_{ij}^{(g)}|\theta] = E \left[ \left( \sum_l H_{il}(\theta|Z) \frac{\Delta_l}{\Delta_j} \right)^2 \right] - \left( E \left[ \sum_l H_{il}(\theta|Z) \frac{\Delta_l}{\Delta_j} \right] \right)^2
\]

\[+ O(c^2)\]

\[= \sum_{l_1, l_2} E \left[ \{H_{il_1}(\theta|Z)H_{il_2}(\theta|Z)\} \right] \frac{\Delta_{l_1} \Delta_{l_2}}{\Delta_j^2} - \left( E \left[ H_{ij}(\theta|Z) \right] \right)^2 \]

\[+ O(c^2)\]

\[= \sum_l \var[H_{il}(\theta|Z)|\theta] + (E[H_{il}(\theta|Z)|\theta])^2 \frac{\Delta_l^2}{\Delta_j^2}
\]

\[= \sum_l b_l(j) \var[H_{il}(\theta|Z)|\theta] + \sum_{l \neq j} b_l(j) \left( E[H_{il}(\theta|Z)|\theta] \right)^2 + O(c^2),(2.41)\]

in which \( b_l(j) = E[\Delta_l^2/\Delta_j^2] \). The first term after the second equality follows from the fact that \( \Delta \) and \( Z \) are statistically independent of each other and the second term has already been seen in the derivation of \( E[\hat{J}_{ij}^{(g)}|\theta] \) in (2.37). The third equality follows by C.1 and I.

By using an identical argument to the one that is mentioned right after (2.40), the expression of \( \text{cov}[\hat{J}_{ij}^{(g)}, \hat{J}_{ji}^{(g)}|\theta], j \neq i \), can be readily shown to be given by,

\[
\text{cov}[\hat{J}_{ij}^{(g)}, \hat{J}_{ji}^{(g)}|\theta] = \text{cov} \left[ \left( \sum_l H_{il}(\theta|Z) \frac{\Delta_l}{\Delta_j} \right), \left( \sum_l H_{jl}(\theta|Z) \frac{\Delta_l}{\Delta_i} \right) \right] + O(c^2). (2.42)
\]
Simplification of the first term in the right-hand-side of (2.42) results in,

\[ \text{cov} \left[ \left( \sum_i H_{ii}(\theta|Z) \frac{\Delta_i}{\Delta_j} \right), \left( \sum_l H_{jl}(\theta|Z) \frac{\Delta_l}{\Delta_j} \right) \right] = E \left[ \left( \sum_i H_{ii}(\theta|Z) \frac{\Delta_i}{\Delta_j} \right) \left( \sum_l H_{jl}(\theta|Z) \frac{\Delta_l}{\Delta_j} \right) \right] - E \left[ H_{ij}(\theta|Z) \right] E \left[ H_{ji}(\theta|Z) \right] - (F_{ij}(\theta))^2 \]

\[ \text{cov} \left[ \hat{J}_{ij}^{(g)}, \hat{J}_{ji}^{(g)} \right] = \left\{ \text{var} \left[ H_{ij}(\theta|Z) \right] + \left( E \left[ H_{ij}(\theta|Z) \right] \right)^2 \right\} + E \left[ H_{ii}(\theta|Z) H_{jj}(\theta|Z) \right] - (F_{ij}(\theta))^2 + O(c^2), \quad j \neq i. \quad (2.44) \]

Now, the variance of \( \hat{H}_{ij}^{(g)} \), \( \text{var}[\hat{H}_{ij}^{(g)}|\theta] \), for \( j \neq i \), can be readily obtained from (2.39) by using (2.41) and (2.44). Note that \( \text{var}[\hat{H}_{ii}^{(g)}|\theta] \) is same as \( \text{var}[\hat{J}_{ii}^{(g)}|\theta] \) that can be directly obtained from (2.41) by replacing \( j \) with \( i \). In the next section, the contributions of the variance and covariance terms to \( \text{var}[\hat{H}_{ij}^{(g)}|\theta] \) would be compared.
with the contributions of the respective variance and covariance terms (as appeared in (2.39)) to \( \text{var}[\hat{\theta}_{ij}^{(g)}|\theta] \) to prove that \( \text{var}[\tilde{\theta}_{ij}^{(g)}|\theta] \) is less than \( \text{var}[\hat{\theta}_{ij}^{(g)}|\theta] \).

Mean and Variance of \( \tilde{\theta}_{ij}^{(g)} \)

It is convenient to express \( \tilde{J}^{(g)}_{ij} \) as below,

\[
\tilde{J}^{(g)}_{ij} = \hat{J}^{(g)}_{ij} - \sum_{l \in I_i} (-F_{il}(\theta)) \frac{\Delta_i}{\Delta_j}, \quad \forall j \in I^c_i,
\]

where it can be readily shown that the \((i, j)\)-th element of \( \tilde{H}_k \) associated with (2.4) is given by,

\[
\tilde{H}^{(g)}_{ij} = \hat{H}^{(g)}_{ij} - \frac{1}{2} \left[ \sum_{l \in I_i} (-F_{il}(\theta)) \frac{\Delta_i}{\Delta_j} + \sum_{l \in I_j} (-F_{jl}(\theta)) \frac{\Delta_j}{\Delta_i} \right], \quad \forall j \in I^c_i
\]

and

\[
\tilde{H}^{(g)}_{ij} = -F_{ij}(\theta), \quad \forall j \in I_i.
\]

As shown for \( \hat{H}^{(L)}_{ij} \) earlier in Case 1, it can shown in a similar way that \( \forall i = 1, \ldots, p, \)

\[
E[\tilde{H}^{(g)}_{ij}|\theta] = \begin{cases} 
-F_{ij}(\theta) + O(c^2), & \forall j \in I^c_i, \\
-F_{ij}(\theta), & \forall j \in I_i.
\end{cases}
\]

Again, the expressions of the ‘big-O’ terms both in (2.38) and in the first equation of (2.48) are precisely same implying that \( E[\tilde{H}^{(g)}_{ij}|\theta] = E[\hat{H}^{(g)}_{ij}|\theta], \forall j \in I^c_i. \)
Next to compute \( \text{var}[\tilde{H}_{ij}(\theta)] \), \( \forall j \in \mathbb{I}_i \), \( \tilde{J}_{ij}^{(g)} \) in (2.36) is decomposed into two parts in a similar fashion as \( \tilde{J}_{ij}^{(L)} \) was decomposed for Case 1. Subsequent approximation of the unknown elements of \( \mathbf{H}(\theta|\mathbf{Z}) \) (particularly the ones that correspond to the elements of the FIM that are known \( a \) priori) in the resulting expression of \( \tilde{J}_{ij}^{(g)} \) by the negative of those elements of the FIM yields the following,

\[
\tilde{J}_{ij}^{(g)} = \tilde{J}_{ij}^{(g)} - \sum_{l \in \mathbb{I}_i} (-F_{il}(\theta)) \frac{\Delta_l}{\Delta_j} = \sum_l X_{il}(\theta|\mathbf{Z}) \frac{\Delta_l}{\Delta_j} + O_{\Delta\mathbf{Z}}(c^2) \quad \forall j \in \mathbb{I}_i, \tag{2.49}
\]

in which the equality follows, as just described, by the use of (2.24) and (2.28). Therefore, by the similarity between (2.36) and (2.49) and by the use of (2.28) and (2.25)-(2.26), \( \text{var}[\tilde{J}_{ij}^{(g)}|\theta], \forall j \in \mathbb{I}_i, \forall i = 1, \cdots, p \), follows readily from (2.41) by replacing \( H_{lm} \) with \( X_{lm} \) as reported below,

\[
\text{var}[\tilde{J}_{ij}^{(g)}|\theta] = \sum_l b_l(j) \text{var} [H_{il}(\theta|\mathbf{Z})|\theta] + \sum_{l \in \mathbb{I}_i \setminus \{ j \}} b_l(j) (E [H_{il}(\theta|\mathbf{Z})|\theta])^2 + O(c^2). \tag{2.50}
\]

Subtracting (2.50) from (2.41), it can be readily seen that \( \forall j \in \mathbb{I}_i \),

\[
\text{var}[\tilde{J}_{ij}^{(g)}|\theta] - \text{var}[\tilde{J}_{ij}^{(g)}|\theta] = \sum_{l \in \mathbb{I}_i} b_l(j) (E [H_{il}(\theta|\mathbf{Z})|\theta])^2 + O(c^2) = \sum_{l \in \mathbb{I}_i} b_l(j) (F_{il}(\theta))^2 + O(c^2) > 0. \tag{2.51}
\]

Here, the last inequality follows from the fact that \( b_l(j) = E[\Delta_l^2/\Delta_j^2] > 0, l = 1, \cdots, p \), for any given \((i, j)\) and assuming that at least one of the known elements, \( F_{il}(\theta) \), in (2.51) is not equal to zero. Note that \( b_l(j) \) turns out to be unity if \( \Delta_1, \cdots, \Delta_p \) is assumed to be Bernoulli \( \pm 1 \) i.i.d. random variables and, as always, the bias term,
\( O(c^2) \), can be made negligibly small by selecting the user-controlled variable \( c \) small enough.

As remarked earlier in the context of Case 1 (p. 34), it can be noted that \( \text{var}[\tilde{H}^{(g)}_{ii}|\theta] \) < \( \text{var}[\tilde{H}^{(g)}_{ii}|\theta] \) if \( j = i \in \mathbb{I}_c \), since \( \text{var}[\tilde{H}^{(g)}_{ii}|\theta] \) and \( \text{var}[\tilde{H}^{(g)}_{ii}|\theta] \) are precisely and respectively given by \( \text{var}[\tilde{j}^{(g)}_{ii}|\theta] \) and \( \text{var}[\tilde{j}^{(g)}_{ii}|\theta] \) and, \( \text{var}[\tilde{j}^{(g)}_{ii}|\theta] < \text{var}[\tilde{j}^{(g)}_{ii}|\theta] \) by (2.51).

Now, since variance of \( \tilde{H}^{(g)}_{ij} \) in (2.46) is given by,

\[
\text{var}[\tilde{H}^{(g)}_{ij}|\theta] = \frac{1}{4} \left( \text{var}[\tilde{j}^{(g)}_{ij}|\theta] + \text{var}[\tilde{j}^{(g)}_{ji}|\theta] + 2\text{cov}[\tilde{j}^{(g)}_{ij}, \tilde{j}^{(g)}_{ji}|\theta] \right),
\]

the covariance term, \( \text{cov}[\tilde{j}^{(g)}_{ij}, \tilde{j}^{(g)}_{ji}|\theta], j \neq i, \forall j \in \mathbb{I}_c \), is computed and compared to \( \text{cov}[\tilde{j}^{(g)}_{ij}, \tilde{j}^{(g)}_{ji}|\theta] \) next to conclude that \( \text{var}[\tilde{H}^{(g)}_{ij}|\theta] < \text{var}[\tilde{H}^{(g)}_{ij}|\theta], \forall j \in \mathbb{I}_c \). The covariance term, \( \text{cov}[\tilde{j}^{(g)}_{ij}, \tilde{j}^{(g)}_{ji}|\theta], j \neq i, \forall j \in \mathbb{I}_c \), follows readily by substituting \( H_{lm} \) with \( X_{lm} \) and using the identical arguments employed in deducing (2.42) and (2.43) yielding a similar expression to (2.44), for \( j \neq i, \forall j \in \mathbb{I}_c, i = 1, \cdots, p, \)

\[
\text{cov}[\tilde{j}^{(g)}_{ij}, \tilde{j}^{(g)}_{ji}|\theta] = \{ \text{var}[(X_{ij}(\theta|Z))|\theta] + (E[(X_{ij}(\theta|Z))|\theta])^2 \}
+ E[X_{ij}(\theta|Z)X_{jj}(\theta|Z)|\theta] - (F_{ij}(\theta))^2 + O(c^2)
= \{ \text{var}[(H_{ij}(\theta|Z))|\theta] + (E[(H_{ij}(\theta|Z))|\theta])^2 \}
+ E[X_{ii}(\theta|Z)X_{jj}(\theta|Z)|\theta] - (F_{ij}(\theta))^2 + O(c^2).
\]

The second equality follows from the definition of \( X_{ij} \) in (2.28) by particularly noting the fact that \( j \in \mathbb{I}_c \). As already indicated in the context of \( \text{cov}[\tilde{j}^{(L)}_{ij}, \tilde{j}^{(L)}_{ji}|\theta] \) for Case 1 (see after (2.33), p. 35), \( X_{ii} \) and \( X_{jj} \) must take the form from one of following four possibilities: (1) \( e_{ii} \) and \( e_{jj} \), (2) \( e_{ii} \) and \( H_{jj} \), (3) \( H_{ii} \) and \( e_{jj} \), (4) \( H_{ii} \) and \( H_{jj} \). It is also
shown there that \( E[H_{ii}(\theta|Z)H_{jj}(\theta|Z)|\theta] - E[X_{ii}(\theta|Z)X_{jj}(\theta|Z)|\theta] \) is \( F_{ii}(\theta)F_{jj}(\theta) > 0 \) for the first, second and third possibilities and 0 for the fourth possibility. Since the difference between \( \text{cov}[\hat{J}_{ij}^{(g)}, \hat{J}_{ji}^{(g)}|\theta] \) in (2.44) and \( \text{cov}[\tilde{J}_{ij}^{(g)}, \tilde{J}_{ji}^{(g)}|\theta] \) in (2.53) is given by,

\[
\text{cov}[\hat{J}_{ij}^{(g)}, \hat{J}_{ji}^{(g)}|\theta] - \text{cov}[\tilde{J}_{ij}^{(g)}, \tilde{J}_{ji}^{(g)}|\theta] = E[H_{ii}(\theta|Z)H_{jj}(\theta|Z)|\theta] - E[X_{ii}(\theta|Z)X_{jj}(\theta|Z)|\theta] + O(c^2), \text{ for } j \neq i, \forall j \in \mathbb{I}_i,(2.54)
\]

it can be immediately concluded from (2.39) and (2.52) that

\[
\text{var}[\tilde{H}_{ij}^{(g)}|\theta] < \text{var}[\hat{H}_{ij}^{(g)}|\theta], \quad i, j = 1, \ldots, p,
\]

by using (2.51) and the fact that \( E[H_{ii}(\theta|Z)H_{jj}(\theta|Z)|\theta] - E[X_{ii}(\theta|Z)X_{jj}(\theta|Z)|\theta] \geq 0 \), that appeared in (2.54), and also noting that the bias term, \( O(c^2) \), can be made negligibly small by selecting \( c \) small enough. Note that the ranges of \( i \) and \( j \) are mentioned here as \( \{1, \ldots, p\} \) not just \( j \in \mathbb{I}_i \). This is simply because for the other case, \( j \in \mathbb{I}_i \), \( \text{var}[\tilde{H}_{ij}^{(g)}|\theta] = 0 \), by (2.47), and therefore, \( \text{var}[\tilde{H}_{ij}^{(g)}|\theta] < \text{var}[\hat{H}_{ij}^{(g)}|\theta], \quad j \in \mathbb{I}_i \).

As a special remark it should be observed here that the several derivations for Case 2 — the case when direct measurements of \( g \) are available — as described in this section remain still valid even if the condition on \( H_{lm,s} \) in C.4 is relaxed.

Finally, since \( \tilde{H}_k, k = 1, \ldots, N \), are statistically independent of each other and \( \hat{H}_k, k = 1, \ldots, N \), are also statistically independent of each other, it can be concluded straightway that \((i, j)\)-th element of \( \tilde{F}_n = -(1/N)\sum_{k=1}^N \tilde{H}_k \) and \( \hat{F}_n = -(1/N)\sum_{k=1}^N \hat{H}_k \)
(with appropriate superscript, $(L)$ or $(g)$) satisfy the following relation,

$$\text{var}[\tilde{F}_{ij}|\theta] = \frac{\text{var}[\tilde{H}_{ij}|\theta]}{N} < \text{var}[\hat{F}_{ij}|\theta] = \frac{\text{var}[\hat{H}_{ij}|\theta]}{N}, \quad i, j = 1, \cdots, p. \quad (2.55)$$

Therefore, we conclude this section by stating that the better estimator, $\tilde{F}_n$, (vis-à-vis the current estimator, $\hat{F}_n$) as determined by using the modified resampling algorithm would preserve the exact analytically known elements of $F_n(\theta)$ as well as reduce the variances of the estimators of the unknown elements of $F_n(\theta)$.

Next chapter presents a few examples illustrating the effectiveness of the modified resampling algorithm.
Chapter 3

Numerical Illustrations and Discussions

In this chapter, two example problems are presented. A problem with a scalar-valued random variable is considered in the first example and its multivariate version is considered in the second example.

3.1 Example 1

Consider independently distributed scalar-valued random data $z_i$ with $z_i \sim N(\mu, \sigma^2 + c_i \alpha)$, $i = 1, \cdots, n$, in which $\mu$ and $(\sigma^2 + c_i \alpha)$ are, respectively, mean and variance of $z_i$ with $c_i$ being some known nonnegative constants and $\alpha > 0$. Here, $\theta$ is considered as $\theta = [\mu, \sigma^2, \alpha]^{T}$. This is a simple extension of an example problem already considered in literature [12, Example 13.7]. The analytical FIM, $\mathbf{F}_n(\theta)$, can be readily
determined for this case so that the MC resampling-based estimates of $F_n(\theta)$ can be verified with the analytical FIM. It can be shown that the analytical FIM is given by,

$$
F_n(\theta) = \begin{bmatrix}
F_{11} & 0 & 0 \\
0 & F_{22} & F_{33} \\
0 & F_{33} & F_{33}
\end{bmatrix},
$$

in which $F_{11} = \sum_{i=1}^{n}(\sigma^2 + c_i \alpha)^{-1}$, $F_{22} = (1/2)\sum_{i=1}^{n}(\sigma^2 + c_i \alpha)^{-2}$ and $F_{33} = (1/2)\sum_{i=1}^{n} c_i (\sigma^2 + c_i \alpha)^{-2}$. Here, the value of $\theta$, that is used to generate the pseudo data vector (as a proxy for $Z_n = [z_1, \cdots, z_n]^T$) and to evaluate $F_n(\theta)$, is assumed to correspond to $\mu = 0$, $\sigma^2 = 1$ and $\alpha = 1$. The values of $c_i$ across $i$ are chosen between 0 and 1 which are generated by using MATLAB uniform random number generator, `rand`, with a given seed (`rand('state',0)`). Based on $n = 30$ yields a positive definite $F_n(\theta)$ whose eigenvalues are given by 0.5696, 8.6925 and 20.7496.

To illustrate the effectiveness of the modified MC based resampling algorithm, it is assumed here that only the upper-left $2 \times 2$ block of the analytical FIM is known a priori. Using this known information, both the existing [13] and the modified resampling algorithm as proposed in this work are employed to estimate the FIM. The number of pseudo data vectors generated for both the algorithms is considered as $N = 2000$. For Hessian estimation per (1.2), $c$ is considered as 0.0001 and, for gradient-approximation per (1.3), $\tilde{c}$ is considered as 0.00011. Bernoulli ±1 random variable components are considered to generate both the perturbation vectors, $\Delta_k$ and $\tilde{\Delta}_k$. 

47
Discussions on results are presented after the description of Example 2.

3.2 Example 2

Consider independently distributed random data vector $\mathbf{Z}_i$ with $\mathbf{Z}_i \sim N(\mu, \Sigma + \mathbf{P}_i)$, $i = 1, \ldots, n$, in which $\mu$ and $(\Sigma + \mathbf{P}_i)$ are, respectively, mean vector and covariance matrix of $\mathbf{Z}_i$ with $\mathbf{P}_i$ being a known matrix such that $(\Sigma + \mathbf{P}_i)$ is a positive definite matrix. Here, $\theta$ is considered as the elements of $\mu$ and the functionally independent elements of $\Sigma$ (i.e., elements of the upper or lower part of the symmetric matrix $\Sigma$ including the diagonal elements). Hence, if $\mathbf{Z}_i$ is a $M$-dimensional random data vector and, $\mu$ is given by $\mu = [\mu_1, \ldots, \mu_M]^T$ and $(i, j)$-th element of $\Sigma$ is given by $\Sigma_{ij}$, $i, j = 1, \ldots, M$, then $\theta$ is considered as $\theta = [\mu_1, \ldots, \mu_M, \Sigma_{11}, \Sigma_{12}, \Sigma_{22}, \ldots, \Sigma_{1M}, \ldots, \Sigma_{MM}]^T$ with $p = \dim(\theta) = M + M(M + 1)/2$. This is a multivariate version of the previous example and have already been considered in literature [13].

It can be readily shown that the log-likelihood function, $L$, for this example is given by,

$$L(\theta|\mathbf{Z}_n) = -\frac{nM}{2} \ln(2\pi) - \frac{1}{2} \sum_i \ln \det (\Sigma + \mathbf{P}_i)$$

$$-\frac{1}{2} \sum_{i=1}^n (\mathbf{Z}_i - \mu)^T (\Sigma + \mathbf{P}_i)^{-1} (\mathbf{Z}_i - \mu).$$

Now, using a simplified notation for derivative, let us denote the derivative of a matrix
$C(\theta)$, whose elements depend on $\theta$, with respect to the $l$-th element, $\theta_l$, of $\theta$ by,

$$C(\theta)_l \equiv \frac{\partial C(\theta)}{\partial \theta_l}.$$ 

Then, the $l$-th element of the gradient vector, $g$, is given by [15], $l = 1, \cdots, p$,

$$g_l(\theta|Z_n) = -\frac{1}{2} \sum_{i=1}^{n} \text{tr} \left[ (\Sigma + P_i)^{-1} (\Sigma + P_i)_l - (\Sigma + P_i)^{-1} (\Sigma + P_i)_l (\Sigma + P_i)^{-1} \right.$$ 

$$\left. \left\{ (Z_i - \mu)(Z_i - \mu)^T \right\} + (\Sigma + P_i)^{-1} \left\{ (Z_i - \mu)(Z_i - \mu)^T \right\}_l \right],$$

in which $\text{tr}$ represents the trace of a matrix. Finally, the $p \times p$ analytical FIM, $F_n(\theta)$, can also be determined and can be compared to the MC resampling-based estimates of $F_n(\theta)$. The $(l, m)$-th element of the analytical FIM, $F_n(\theta)$, is given by [15], $l, m = 1, \cdots, p$,

$$F_{lm}(\theta) = \frac{1}{2} \sum_{i=1}^{n} \text{tr} \left[ A^{(i)}_l A^{(i)}_m + (\Sigma + P_i)^{-1} M_{lm} \right], \quad (3.1)$$

in which $A^{(i)}_l \equiv (\Sigma + P_i)^{-1} (\Sigma + P_i)_l$ and $M_{lm} \equiv \mu_i \mu_m^T + \mu_m \mu_i^T$.

In the current example, the dimension of the random data vector, $Z_i$, is considered to be $M = 4$ so that $p = 4 + 4(4 + 1)/2 = 14$. The value of $\theta$, that is used to generate the pseudo data vector (as a proxy for $Z_n = [Z_1, \cdots, Z_n]^T$) and to evaluate $F_n(\theta)$, corresponds to $\mu = [0, 0, 0, 0]^T$, $\Sigma$ being a $4 \times 4$ matrix with 1 on the diagonal and 0.5 on the off-diagonals and $P_i$ being a $4 \times 4$ matrix of zero elements for all $i$. Using $n = 30$, (3.1) then yields a positive definite $F_n(\theta)$ with smallest eigenvalue 3.8207 and largest eigenvalue 120.0000.

To illustrate the effectiveness of the modified MC based resampling algorithm for this example, it is assumed here that only the upper-left $4 \times 4$ block of the FIM
associated with the elements of $\mu$ and, all the diagonal elements of the rest of FIM (associated with the functionally independent elements of $\Sigma$, i.e., associated with the elements of the upper triangular part of $\Sigma$) are known \textit{a priori}. The structure of the FIM in terms of the assumed known elements is shown schematically in Figure 3.1. Using this known information, both the existing [13] and the modified resampling algorithm as proposed in this work are employed to estimate the FIM.

The number of pseudo data vectors generated for both the algorithms is considered as $N = 40000$. For Hessian estimation per (1.2), $c$ is considered as 0.0001 and, for gradient-approximation per (1.3), $\tilde{c}$ is considered as 0.00011. Bernoulli $\pm 1$ random variable components are considered to generate both the perturbation vectors, $\Delta_k$ and $\tilde{\Delta}_k$.

It should be remarked here that the benefit of the recursive representation of sample mean of the Hessian estimates as mentioned at p. 9 could not be realized for...
Example 1 since the computation cost involved in its FIM estimation was negligible. However, use of the standard recursive representation of sample mean of the (random) elements of $\hat{H}_k$ (with appropriate superscripts, $(L)$ or $(g)$) have enormously reduced the computation time for this example problem. Though in practice computation of sample variance is usually not required, sample variance of the (random) elements of $\hat{H}_k$ are computed to compare with that of $\tilde{H}_k$ in order to investigate the effectiveness of the modified resampling algorithm. If $\text{var}[H(\theta|Z_n)|\theta]$ represents a matrix whose $(l,m)$-th element is given by $\text{var}[H_{lm}(\theta|Z_n)|\theta]$, then the recursive representation of estimate of $\text{var}[H(\theta|Z_n)|\theta]$ is given by,

$$\text{var}_{\hat{H}_k} = \frac{k-1}{k} \text{var}_{\hat{H}_{k-1}} + \frac{1}{k-1} \left\{ \left( \hat{H}_k - \bar{H}_k \right) \circ \left( \hat{H}_k - \bar{H}_k \right) \right\}, \quad k = 2, \ldots, N, $$

(3.2)
in which at the first step ($k = 1$), $\text{var}_{\hat{H}_1}$ in the left-hand-side is set to a matrix of zeros (i.e., the variance computation essentially starts from $k = 2$), the symbol $\circ$ represents the element-wise product operator (also known as Hadamard product operator) and $\bar{H}_k$ is computed by using (1.4). At the end of the $N$-th step, $\text{var}_{\tilde{H}_N}$ is normalized by $(N-1)$ by setting $\text{var}_{\tilde{H}_N} = (N/(N-1))\text{var}_{\tilde{H}_N}$ to obtain the unbiased estimator of $\text{var}[H(\theta|Z_n)|\theta]$. An identical expression is also used for $\bar{H}_k$. Note that the $(l,m)$-th element of $\text{var}_{\hat{H}_k}$ and $\text{var}_{\tilde{H}_k}$ are, respectively, same as $\text{var}[\hat{H}_{lm}|\theta]$ and $\text{var}[\tilde{H}_{lm}|\theta]$ (see (2.55)).

Below the discussions on results for both the examples as described above are presented.
3.3 Discussions on results

The FIM is estimated based on both the log-likelihood measurements (Case 1) and the gradient vector measurements (Case 2). The results for Example 1 and Example 2 are, respectively, tabulated in Table 3.1 and Table 3.3. The absolute mean-squared error (MSE) of $\hat{F}_n$ and $\tilde{F}_n$ as presented in the third and fifth columns of these tables are computed, for example, in the case of $\hat{F}_n$ in the third column, as $\text{MSE}(\hat{F}_n) = \sum_{ij} (\hat{F}_{ij} - F_{ij}(\theta))^2$. The relative MSE as presented in the second and fourth columns are computed, for example, in the case of $\hat{F}_n$, as $\text{relMSE}(\hat{F}_n) = 100 \times \frac{\text{MSE}(\hat{F}_n)}{\sum_{ij} (F_{ij}(\theta))^2}$. The effectiveness of the modified resampling algorithm can be clearly seen from the sixth column of these tables that show substantial MSE reduction. The relative MSE reduction in these tables are computed as $100 \times \frac{(\text{MSE}(\hat{F}_n) - \text{MSE}(\tilde{F}_n))}{\text{MSE}(\hat{F}_n)}$. In these columns also shown within parentheses are variance reduction. The relative variance reduction are computed as $100 \times \frac{(A - B)}{A}$, in which $A = \sum_{ij} \text{var}[\hat{F}_{ij}\mid \theta]$ and $B = \sum_{ij} \text{var}[\tilde{F}_{ij}\mid \theta]$.

It would also be interesting to investigate the effect of the modified resampling algorithm on the MSE reduction in the estimators of the unknown elements of the FIM in contrast to a rather ‘naive approach’ in which the estimates of the unknown elements are simply extracted from $\tilde{F}_n$. To see the improvement in terms of MSE reduction of the estimators of the unknown elements of the FIM, the elements corresponding to the upper-left $2 \times 2$ block for Example 1 and the elements as shown in Figure 3.1 for Example 2 of $\tilde{F}_n$ obtained from the current resampling algorithm are
replaced by the corresponding known analytical elements of the FIM, $F_n(\theta)$. Clearly, the contribution of MSE of the elements of $\hat{F}_n$ corresponding to the known elements of the FIM are now zero to $\text{MSE}(\hat{F}_n)$ like in the context of $\tilde{F}_n$ in which the contribution of MSE of the elements of $\tilde{F}_n$ corresponding to the known elements of the FIM are zero to $\text{MSE}(\tilde{F}_n)$. Therefore, the results as shown in Table 3.2 and Table 3.4 only display the contributions of the MSE from the estimators of the unknown elements of FIM. The relative MSE reduction in the last columns of these tables are reported as earlier by showing $100 \times (\text{MSE}(\hat{F}_n) - \text{MSE}(\tilde{F}_n))/\text{MSE}(\hat{F}_n)$. These tables clearly reflect the superiority of the modified resampling algorithm as presented in this work over the current resampling algorithm. In these tables, similar results on variance are also reported within parentheses.

Let us now focus more carefully on the numerical values shown in these tables. For example, consider Case 2 in Table 3.2 in which the MSE reduction of 0.7930 % is much less than the variance reduction of 94.4222 %. It is known [12, Section 13.1.1] that the MSE can be decomposed into two parts – one due to the variance and the other being contributed from bias. The fact that the 94.4222 % variance reduction is associated with a very small MSE reduction of 0.7930 % (from 0.0885 to 0.0878 only) clearly indicates that the effect of bias to the MSE is relatively much more than the effect of variance to the MSE. Notwithstanding the fact that this significant observation is against the backdrop of very small MSE relative to the true FIM with $\text{relMSE}(\hat{F}_n) = 0.0175$ % and $\text{relMSE}(\tilde{F}_n) = 0.0173$ %, it worths a mention. The
Table 3.1: Example 1 – MSE and MSE reduction of FIM estimates based on \( N = 2000 \) pseudo data (results on variance are reported within parentheses).

<table>
<thead>
<tr>
<th>Cases</th>
<th>Error in FIM estimates</th>
<th>MSE (variance) reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>relMSE(( \hat{F}_n ))</td>
<td>MSE(( \hat{F}_n ))</td>
</tr>
<tr>
<td>Case 1</td>
<td>0.3815 %</td>
<td>1.9318</td>
</tr>
<tr>
<td>Case 2</td>
<td>0.0533 %</td>
<td>0.2703</td>
</tr>
</tbody>
</table>

other numbers in Table 3.2 and Table 3.4 can be interpreted in a similar fashion.

Table 3.1 to Table 3.4 essentially highlight the substantial improvement of the results (in the sense of MSE reduction as well as variance reduction) of the modified MC based resampling algorithm over the results of the current MC based resampling algorithm. Of course, this degree of improvement is controlled by the values of the known elements of the analytical FIM; see (2.32) and (2.34) for Case 1 and (2.51) and (2.54) for Case 2. Finally, to validate the correctness of simulation results as presented above, theoretical difference between the variance of estimators of the unknown elements of the FIM for Example 1 as computed by using (2.32), (2.34), (2.51), (2.54), (2.13), (2.30), (2.39), (2.52), and (2.55), are presented in Table 3.5 against the same quantities obtained from simulation.
\[
\text{MSE}(\hat{F}_n) \quad \text{(and } A) \quad \text{MSE}(\tilde{F}_n) \quad \text{(and } B) \quad \text{MSE (variance)}
\]

<table>
<thead>
<tr>
<th>Cases</th>
<th>MSE(\hat{F}_n)</th>
<th>MSE(\tilde{F}_n)</th>
<th>MSE(\hat{F}_n)</th>
<th>MSE(\tilde{F}_n)</th>
<th>relMSE(\hat{F}_n)</th>
<th>MSE(\tilde{F}_n)</th>
<th>relMSE(\tilde{F}_n)</th>
<th>reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>0.1288 (0.0159)</td>
<td>0.1021 (0.0006)</td>
<td>20.7235 %</td>
<td>95.9179 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 2</td>
<td>0.0885 (0.0030)</td>
<td>0.0878 (0.0002)</td>
<td>0.7930 %</td>
<td>94.4222 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2: Example 1 – MSE comparison for \( \hat{F}_n \) and \( \tilde{F}_n \) for only the unknown elements of \( F_n(\theta) \) according to \( F_n^{(\text{given})} \) based on \( N = 100000 \) pseudo data (similar results on variance are reported within parentheses, \( A = \sum_{i=1}^{p} \sum_{j \in I_i} \text{var}[\hat{F}_{ij}|\theta] \) and \( B = \sum_{i=1}^{p} \sum_{j \in I_i} \text{var}[\tilde{F}_{ij}|\theta] \)).

\[
\text{MSE (variance)}
\]

<table>
<thead>
<tr>
<th>Cases</th>
<th>relMSE(\hat{F}_n)</th>
<th>MSE(\hat{F}_n)</th>
<th>relMSE(\tilde{F}_n)</th>
<th>MSE(\tilde{F}_n)</th>
<th>MSE (variance) reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>0.5991 %</td>
<td>430.1066</td>
<td>0.1516 %</td>
<td>108.8217</td>
<td>74.6989 % (84.5615 %)</td>
</tr>
<tr>
<td>Case 2</td>
<td>0.3860 %</td>
<td>277.1268</td>
<td>0.0966 %</td>
<td>69.3549</td>
<td>74.9736 % (88.7458 %)</td>
</tr>
</tbody>
</table>

Table 3.3: Example 2 – MSE and MSE reduction of FIM estimates based on \( N = 40000 \) pseudo data (results on variance are reported within parentheses).

\[
\text{MSE (variance)}
\]

<table>
<thead>
<tr>
<th>Cases</th>
<th>MSE(\hat{F}_n)</th>
<th>MSE(\tilde{F}_n)</th>
<th>MSE(\hat{F}_n)</th>
<th>MSE(\tilde{F}_n)</th>
<th>relMSE(\hat{F}_n)</th>
<th>MSE(\tilde{F}_n)</th>
<th>relMSE(\tilde{F}_n)</th>
<th>reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>233.3016 (150.0580)</td>
<td>108.8217 (28.0297)</td>
<td>53.3558 %</td>
<td>81.3208 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 2</td>
<td>78.8887 (16.7351)</td>
<td>69.3549 (2.0077)</td>
<td>12.0851 %</td>
<td>88.0030 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.4: Example 2 – MSE comparison for \( \hat{F}_n \) and \( \tilde{F}_n \) for only the unknown elements of \( F_n(\theta) \) according to \( F_n^{(\text{given})} \) based on \( N = 40000 \) pseudo data (similar results on variance are reported within parentheses \( A = \sum_{i=1}^{p} \sum_{j \in I_i} \text{var}[\hat{F}_{ij}|\theta] \) and \( B = \sum_{i=1}^{p} \sum_{j \in I_i} \text{var}[\tilde{F}_{ij}|\theta] \)).
Cases | $\text{var}[F_{13}|\theta] - \text{var}[\tilde{F}_{13}|\theta]$ | $\text{var}[F_{23}|\theta] - \text{var}[\tilde{F}_{23}|\theta]$ | $\text{var}[F_{33}|\theta] - \text{var}[\tilde{F}_{33}|\theta]$ |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>simulation</td>
<td>theory</td>
<td>simulation</td>
</tr>
<tr>
<td>Case 1</td>
<td>0.1417</td>
<td>0.1399</td>
<td>0.1320</td>
</tr>
<tr>
<td>Case 2</td>
<td>0.0629</td>
<td>0.0629</td>
<td>0.0093</td>
</tr>
</tbody>
</table>

Table 3.5: Example 1 – Variance comparison between theoretical results (correct up to a bias term as discussed in Section 2.3) and simulation results (with $N = 2000$).
Chapter 4

Conclusions and Future Research

Analytical evaluation of the Fisher information matrix (FIM), in the general case, is difficult or next to impossible since it involves computation of first or second order derivative of log-likelihood function and expectation of a resulting multivariate and non-linear function. To avoid this formidable task, a simulation-based technique, resampling algorithm [12, Section 13.3.5], [13], is usually recommended for estimation of the FIM. In some practical applications [4, 5], there arise cases in which some of the elements of the FIM are known a priori and the remaining elements unknown (and need to be estimated). In such situations, a naive approach is usually followed — by simply extracting the estimates of the unknown elements from the full estimate (outcome of the existing resampling algorithm) of the FIM — whereby the information contained in the known elements is not exploited to enhance the statistical qualities of the estimators of the unknown elements.
4.1 Summary of Contributions Made

In the current work, it is assumed that the measurements of the log-likelihood function or the gradient vector of log-likelihood function are available. The present work re-visits the resampling algorithm and computes the variance of the estimator of an arbitrary element of the FIM. A modification in the existing resampling algorithm is proposed simultaneously preserving the known elements of the FIM and improving the statistical characteristics of the estimators of the unknown elements (in the sense of variance reduction) by utilizing the information available from the known elements.

Numerical examples presented in Chapter 3 showed significant improvement of the results (in the sense of MSE reduction as well as variance reduction) of the modified MC based resampling algorithm over the results of the current MC based resampling algorithm.

4.2 Suggestions for Future Research

Based on the study carried out in this work, the following are a few suggestions for further research:

1. A way to employ the running feedback (similar to the work presented in [14]) to the averaging process of the resampling algorithm to further enhance the statistical characteristics of the estimators of the unknown elements of the FIM may be investigated.
2. Though the prior information available in terms of the known elements of the FIM are used in the current work to enhance the statistical characteristics of the estimators of the unknown elements of the FIM, the estimates of the known elements resulted from the resampling algorithm are not used since these estimates of the known elements are simply replaced by the analytically known elements. The issue that need further attention is to see if there is any way to consider these “wasted” information as well to further enhance the statistical characteristics of the estimators of the unknown elements of the FIM.
Bibliography


Vita

Sonjoy Das was born on August 22, 1972 in West Bengal, India. He earned a Bachelor of Engineering degree in Civil Engineering from Bengal Engineering College (Deemed University), West Bengal, India in 1995 and a Masters of Science (Engineering) degree in Civil (Structure) Engineering from Indian Institute of Science, Bangalore, India in 2001. He has a total of almost four years of experience in Civil Engineering industry as an assistant engineer and in Indian Space Research Organisation as a scientist. He started his Doctor of Philosophy program in the Department of Civil Engineering at The Johns Hopkins University, Baltimore, USA in September, 2002 and, later transferred to the University of Southern California, Los Angeles, USA in August, 2005 to continue his work. Currently, he is a candidate for the degree of Doctor of Philosophy in the Department of Civil and Environmental Engineering at the University of Southern California and, also concurrently pursuing his Masters of Science in Engineering program in the Department of Applied Mathematics and Statistics at The Johns Hopkins University.