A General Framework for Model Diagnostics: Diagnostic Measures and Influence Analysis

Hongtu Zhu

Department of Biostatistics and Biomedical Research Imaging Center
The University of North Carolina at Chapel Hill

August 22, 2012

Motivation

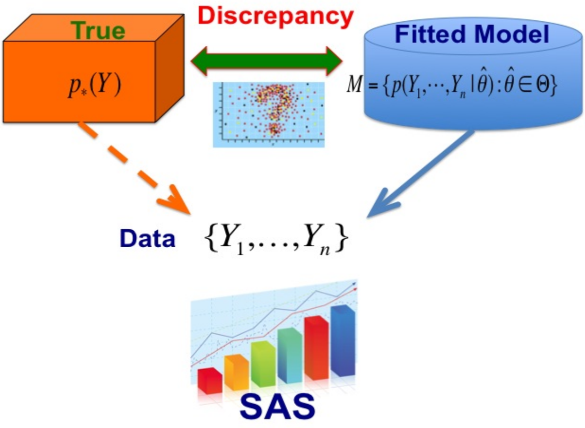

Figure: True complicated process, Data, and 'Right'/'Fitted' model.

Motivation

- Data may come from a true complicated process.
- Finding a 'right'/'fitted' model to interpret a dataset and to approximate the true complicated process.
- Fitted Model $\neq$ True Process
- Discrepancy $=$ Fitted Model$\ominus$True Process
- How do we use statistical tools (or diagnostic measures) to detect such discrepancies?

Motivation

- Discrepancy exists between isolated observations (e.g., influential points and outliers) and the rest of the observations
  - residuals
  - leverages
  - case-deletion measures
- Any systematic discrepancies between the data and the fitted values obtained from statistical models
  - graphical procedures of residuals, such as partial residual and added variable plots
  - goodness-of-fit test statistics and test procedures for testing specific alternatives
  - sensitivity analysis

Motivation

Consider $\mathbf{y} = X\beta + \epsilon$, where $\epsilon = (\epsilon_1, \cdots, \epsilon_n)^T \sim N(\mathbf{0}, \sigma^2 I_n)$.

- The quantity $H = (h_{ij}) = X(X^TX)^{-1}X^T$ is called the hat matrix and h_{ii}, called leverages, can be used for assessing each $\mathbf{x}_i$.
- The raw residuals, $\hat{\mathbf{e}} = (\hat{e}_i) = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - H\mathbf{y} = Q\mathbf{y}$, $Q = I_n - H$, provide important information about the fitted model, such as model misspecification, outliers, and influential points.
- The studentized residual is defined to be $r_i = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$, where $\hat{\sigma}$ is an estimate of σ.
- Cook's distance measures the distance between $\hat{\beta}$ and the estimate of β without the $i-$th observation, denoted by $\hat{\beta}_{(i)}$.

Motivation

- Most diagnostic measures were originally developed under linear regression models (Cook, 1977; Cook and Weisberg, 1982; Chatterjee and Hadi, 1988).
- Considerable research has been devoted to developing diagnostic measures for generalized linear models and models for survival data (Andersen, 1992, Davison and Tsai, 1992; Wei, 1998; Storer and Crowley, 1985; Therneau, Grambsch, and Fleming, 1990; Lin, Wei, and Ying, 1993).
- Diagnostic measures have been developed for various models for clustered data and models for missing data (Christensen et al., 1992; Preisser and Qaqish, 1996; Banerjee and Frees,1997; Haslett, 1999; Zhu, et al. 2001; Fung, et.al, 2002).

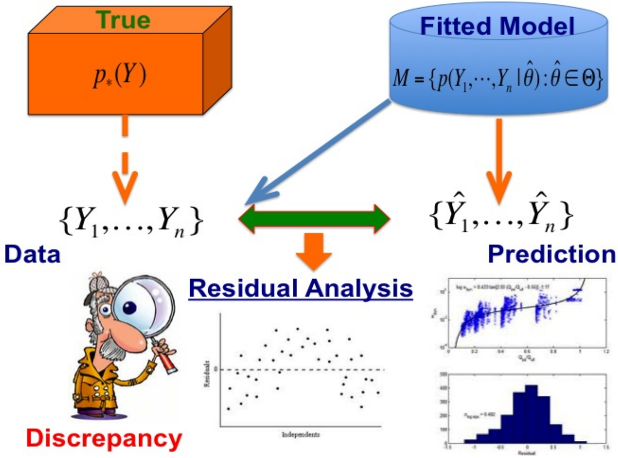

Figure: Residual Analysis

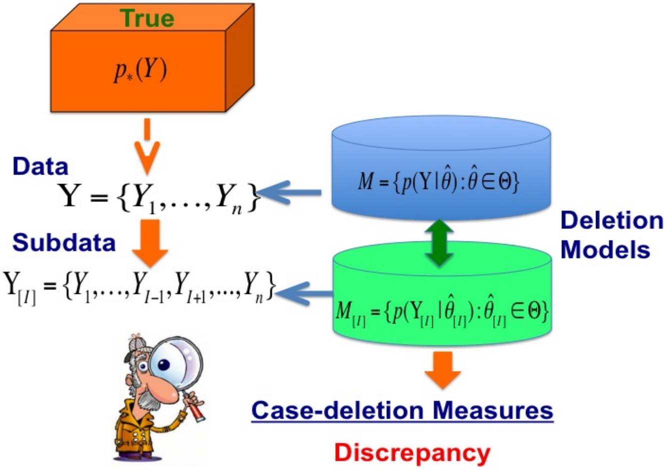

Figure: Case-deletion Perturbation and Measures

Motivation

We address three important issues related to a rigorous method to assess discrepancy among Cook's distance measures:

- the development of a quantity to measure the degree of perturbation introduced by deleting subsets with different numbers of observations;
- the delineation of relationship between the degree of the perturbation and the magnitude of Cook's distance;
- the development of new case-deletion measures for carrying out formal influence analysis.

Cook's Distance

Consider $\mathbf{Y}^T = (Y_1^T, \ldots, Y_n^T)$ and $p(\mathbf{Y}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta \subset R^q$.

- The dimension of $Y_i = (y_{i,1}, \ldots, y_{i,m_i})$ may vary significantly across all i.
- Let subscript '[I]' denote the relevant quantity with all observations in a set I deleted.
- Let $\mathbf{Y}_{[I]}$ be a subsample of $\mathbf{Y}$ with $\mathbf{Y}_I = \{Y_{(i,j)} : (i,j) \in I\}$ deleted and $p(\mathbf{Y}_{[I]}|\boldsymbol{\theta})$ be its probability function.
- $\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}} \log p(\mathbf{Y}|\boldsymbol{\theta})$ and $\hat{\boldsymbol{\theta}}_{[I]} = \text{argmax}_{\boldsymbol{\theta}} \log p(\mathbf{Y}_{[I]}|\boldsymbol{\theta})$;
- $\text{CD}(I) = (\hat{\boldsymbol{\theta}}_{[I]} - \hat{\boldsymbol{\theta}})^T G_{n\theta} (\hat{\boldsymbol{\theta}}_{[I]} - \hat{\boldsymbol{\theta}})$.

Cook's Distance

$$\mathrm{CD}(I) = F_1(I, \mathcal{M}, \mathbf{Y}) = F_2(\mathcal{P}(I|\mathcal{M}), G(I|\mathbf{Y}, \mathcal{M})).$$

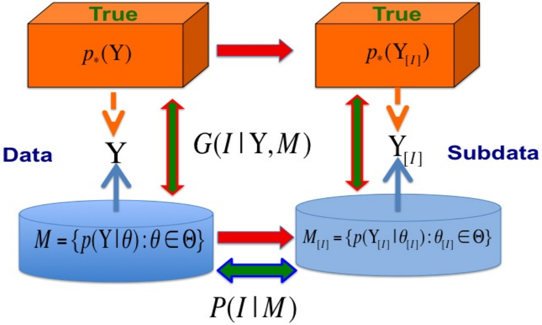

Figure: $G(I|\mathbf{Y}, \mathcal{M})$ is the goodness of fit of $\mathcal{M}$ to $\mathbf{Y}$ for I and $\mathcal{P}(I|\mathcal{M})$ is the degree of the perturbation to $\mathcal{M}$ introduced by deleting the subset I.

Degree of Perturbation

Our choice of $\mathcal{P}(I|\mathcal{M})$ is motivated by five principles as follows.

- (P.a) (non-negativity) For any subset I, $\mathcal{P}(I|\mathcal{M})$ is always non-negative.
- (P.b) (uniqueness) $\mathcal{P}(I|\mathcal{M}) = 0$ if and only if I is an empty set.
- (P.c) (monotonicity) If $I_2 \subset I_1$, then $\mathcal{P}(I_2|\mathcal{M}) \leq \mathcal{P}(I_1|\mathcal{M})$.
- (P.d) (additivity) If $I_2 \subset I_1$, $I_{1\cdot 2} = I_1 - I_2$, and $p(\mathbf{Y}_{I_{1\cdot 2}}|\mathbf{Y}_{[I_1]}, \boldsymbol{\theta}) = p(\mathbf{Y}_{I_{1\cdot 2}}|\mathbf{Y}_{[I_{1\cdot 2}]}, \boldsymbol{\theta})$ for all $\boldsymbol{\theta}$, then we have $\mathcal{P}(I_1|\mathcal{M}) = \mathcal{P}(I_2|\mathcal{M}) + \mathcal{P}(I_{1\cdot 2}|\mathcal{M})$.
- (P.e) $\mathcal{P}(I|\mathcal{M})$ should naturally arise from the current model $\mathcal{M}$, the data $\mathbf{Y}$, and the subset I.

Degree of Perturbation

$\mathcal{P}(I|\mathcal{M})$ is defined as follows.

- $p(\mathbf{Y}|\boldsymbol{\theta}) = p(\mathbf{Y}_{[I]}|\boldsymbol{\theta})p(\mathbf{Y}_I|\mathbf{Y}_{[I]}, \boldsymbol{\theta})$;
- $p(\mathbf{Y}|\boldsymbol{\theta}, I) = p(\mathbf{Y}_{[I]}|\boldsymbol{\theta})p(\mathbf{Y}_I|\mathbf{Y}_{[I]}, \boldsymbol{\theta}_*)$, where $\boldsymbol{\theta}_*$ is the true value of $\boldsymbol{\theta}$ under $\mathcal{M}$;
- $\mathrm{KL}(\mathbf{Y}, \boldsymbol{\theta}|\boldsymbol{\theta}_*, I) = \int p(\mathbf{Y}|\boldsymbol{\theta}) \log\left(\frac{p(\mathbf{Y}|\boldsymbol{\theta})}{p(\mathbf{Y}|\boldsymbol{\theta}, I)}\right) d\mathbf{Y}$;
- $\mathcal{P}(I|\mathcal{M}) = \int \mathrm{KL}(\mathbf{Y}, \boldsymbol{\theta}|\boldsymbol{\theta}_*, I)\phi(\boldsymbol{\theta}|\boldsymbol{\theta}_*, \Sigma_*)d\boldsymbol{\theta}$;
- We suggest substituting $\boldsymbol{\theta}_*$ by an estimator of $\boldsymbol{\theta}$, denoted by $\tilde{\boldsymbol{\theta}}$, and Σ_* by the covariance matrix of $\tilde{\boldsymbol{\theta}}$.

Degree of Perturbation

Theorem 1. *Suppose that* $L(\{\mathbf{Y} : p(\mathbf{Y}_I|\mathbf{Y}_{[I]}, \boldsymbol{\theta}) = p(\mathbf{Y}_I|\mathbf{Y}_{[I]}, \boldsymbol{\theta}_*)\}) > 0$ *for any* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_*$, *where* $L(A)$ *is the Lebesgue measure of a set* A. *Then,* $\mathcal{P}(I|\mathcal{M})$ *defined above satisfies the five principles (P.a)-(P.e).*

Example

Consider the linear regression model $y_i = \mathbf{x}_i^T \boldsymbol{\beta}_* + \epsilon_i$, where $\mathbf{x}_i$ is a $p \times 1$ vector and $\epsilon_i (i.i.d) \sim N(0, \sigma_*^2)$.

- For the case of fixed covariates, $\mathcal{P}(\{i\}|\mathcal{M}) = 0.5E_\theta[\log(\sigma_*^2/\sigma^2)] + 0.5\frac{\mathbf{x}_i^T E_\theta[(\boldsymbol{\beta}-\boldsymbol{\beta}_*)(\boldsymbol{\beta}-\boldsymbol{\beta}_*)^T]\mathbf{x}_i}{\sigma_*^2} \approx \frac{1}{2}h_{ii} + \frac{1}{2n}$.
- For the case of random covariates, we assume that the $\mathbf{x}_i \sim F(\mu_x, \Sigma_x)$ and $\mathcal{P}(\{i\}|\mathcal{M}) = 0.5E_\theta[\log(\sigma_*^2/\sigma^2)] + 0.5\sigma_*^{-2}\text{tr}\{\Sigma_x E_\theta[(\boldsymbol{\beta} - \boldsymbol{\beta}_*)(\boldsymbol{\beta} - \boldsymbol{\beta}_*)^T]\} \approx \frac{p+1}{2n}$.
- $\mathcal{P}(\{i_1, \cdots, i_{n(I)}\}|\mathcal{M}) = \sum_{k=1}^{n(I)} \mathcal{P}(\{i_k\}|\mathcal{M})$.
- For random covariates, we have $\mathcal{P}(I|\mathcal{M}) = n(I)\mathcal{P}(\{1\}|\mathcal{M})$ for any subset I with $n(I)$ observations.
- An important implication of these calculations in real data analysis is that we can directly compare $\text{CD}(I_1)$ and $\text{CD}(I_2)$ when $n(I_1) = n(I_2)$.

Example

- Cook's distance for deleting the subset I with $n(I)$ is given by $CD(I) = \hat{\mathbf{e}}_I^T(\mathbf{I}_{n(I)} - H_I)^{-1}H_I(\mathbf{I}_{n(I)} - H_I)^{-1}\hat{\mathbf{e}}_I/\hat{\sigma}^2$, where $\hat{\mathbf{e}}_I$ is an $n(I) \times 1$ vector containing all $\hat{e}_i$ for $i \in I$ and $H_I = \mathbf{X}_I(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}_I^T$.
- How to compare CD(I_1) and CD(I_2) for any two subsets with $n(I_1) \neq n(I_2)$?

Example

Theorem 2. *For the standard linear model, where* $\mathbf{y} = \mathbf{X}\beta + \epsilon$ *and* $\epsilon \sim N(\mathbf{0}, \sigma^2 I_n)$*, we have the following results:*
(a) for any $I_2 \subset I_1$*,* $CD(I_1)$ *is stochastically larger than* $CD(I_2)$ *for any* $\mathbf{X}$*, that is,* $P(CD(I_1) > t|\mathcal{M}) \geq P(CD(I_2) > t|\mathcal{M})$ *holds for any* $t \geq 0$*.*
(b) Suppose that the components of $\mathbf{X}_I$ *and* $\mathbf{X}_{I'}$ *are identically distributed for any two subsets* I *and* I' *with* $n(I) = n(I')$*. Thus,* $CD(I)$ *and* $CD(I')$ *follow the same distribution when* $n(I) = n(I')$ *and* $CD(I_1)$ *is stochastically larger than* $CD(I_2)$ *for any two subsets* I_2 *and* I_1 *with* $n(I_1) > n(I_2)$*.*

Theorem

Proposition 1. *Under the stochastic larger assumption, for any two subsets I_1 and I_2 with $\mathcal{P}(I_1|\mathcal{M}) > \mathcal{P}(I_2|\mathcal{M})$, Cook's distance satisfies*

$$E[h(CD(I_1))|\mathcal{M}] \quad \geq \quad E[h(CD(I_2))|\mathcal{M}] \tag{1}$$

holds for all increasing functions $h(\cdot)$. In particular, we have $E[CD(I_1)|\mathcal{M}] \geq E[CD(I_2)|\mathcal{M}]$ and $Q_{CD(I_1)}(\alpha|\mathcal{M})$ is greater than the α-quantile of $Q_{CD(I_2)}(\alpha|\mathcal{M})$ for any $\alpha \in [0, 1]$, where $Q_{CD(I)}(\alpha|\mathcal{M})$ denotes the $\alpha-$quantile of the distribution of $CD(I)$ for any subset I.

Definition

Definition 1. The *scaled Cook's distances* for matching (mean, Std) and (median, Mstd) are, respectively, defined as

$$\mathrm{SCD}_1(I) = \frac{\mathrm{CD}(I) - E[\mathrm{CD}(I)|\mathcal{M}]}{\mathrm{Std}[\mathrm{CD}(I)|\mathcal{M}]} \text{ and } \mathrm{SCD}_2(I) = \frac{\mathrm{CD}(I) - Q_{CD(I)}(0.5|\mathcal{M})}{\mathrm{Mstd}[\mathrm{CD}(I)|\mathcal{M}]},$$

where both the expectation and the quantile are taken with respect to $\mathcal{M}$.

Definition 2. The *conditionally scaled Cook's distances* (CSCD) for matching (mean, Std) and (median, Mstd) while controlling for $\mathbf{Z}$ are, respectively, defined as

$$\begin{aligned} \mathrm{CSCD}_1(I, \mathbf{Z}) &= \frac{\mathrm{CD}(I) - E[\mathrm{CD}(I)|\mathcal{M}, \mathbf{Z}]}{\mathrm{Std}[\mathrm{CD}(I)|\mathcal{M}, \mathbf{Z}]}, \\ \mathrm{CSCD}_2(I, \mathbf{Z}) &= \frac{\mathrm{CD}(I) - \mathrm{Q}_{CD(I)}(0.5|\mathcal{M}, \mathbf{Z})}{\mathrm{Mstd}[\mathrm{CD}(I)|\mathcal{M}, \mathbf{Z}]}, \end{aligned}$$

where $\mathbf{Z}$ is the set of some fixed covariates in $\mathbf{Y}$ and the expectation and quantiles are taken with respect to $\mathcal{M}$ given $\mathbf{Z}$.

First-order Approximation

Theorem 3. *If Assumptions A2-A5 in the Appendix hold and $n(I)/n \to \gamma \in [0, 1)$, then we have the following results:*
(a) Let $\mathbf{F}_n(\boldsymbol{\theta}) = -\partial^2_\theta \log p(\mathbf{Y}|\boldsymbol{\theta})$, $\mathbf{f}_I(\boldsymbol{\theta}) = \partial_\theta \log p(\mathbf{Y}_I|\mathbf{Y}_{[I]}, \hat{\boldsymbol{\theta}})$, *and* $\mathbf{s}_I(\boldsymbol{\theta}) = -\partial^2_\theta \log p(\mathbf{Y}_I|\mathbf{Y}_{[I]}, \boldsymbol{\theta})$, *CD(I) can be approximated by*

$$\widetilde{CD}(I) = \mathbf{f}_I(\hat{\boldsymbol{\theta}})^T [\mathbf{F}_n(\hat{\boldsymbol{\theta}}) - \mathbf{s}_I(\hat{\boldsymbol{\theta}})]^{-1} \mathbf{F}_n(\hat{\boldsymbol{\theta}}) [\mathbf{F}_n(\hat{\boldsymbol{\theta}}) - \mathbf{s}_I(\hat{\boldsymbol{\theta}})]^{-1} \mathbf{f}_I(\hat{\boldsymbol{\theta}}); \qquad (2)$$

(b) $E[\widetilde{CD}(I)|\mathcal{M}] \approx tr(\{E[\mathbf{F}_n(\hat{\boldsymbol{\theta}})|\mathcal{M}] - E[\mathbf{s}_I(\hat{\boldsymbol{\theta}})|\mathcal{M}]\}^{-1} E[\mathbf{s}_I(\hat{\boldsymbol{\theta}})|\mathcal{M}])$;
(c) $E[\widetilde{CD}(I)|\mathcal{M}, \mathbf{Z}] \approx$
$tr(\{E[\mathbf{F}_n(\hat{\boldsymbol{\theta}})|\mathcal{M}, \mathbf{Z}] - E[\mathbf{s}_I(\hat{\boldsymbol{\theta}})|\mathcal{M}, \mathbf{Z}]\}^{-1} E[\mathbf{s}_I(\hat{\boldsymbol{\theta}})|\mathcal{M}, \mathbf{Z}])$.

Algorithm

- Step (i). We generate a random sample $\mathbf{Y}^s$ from $p(\mathbf{Y}|\mathbf{Z}, \hat{\theta})$ and calculate $\widetilde{\text{CD}}(I)$ based on the simulated sample $\mathbf{Y}^s$ and fixed $\mathbf{Z}$, denoted by $\widetilde{\text{CD}}(I)^s$.
- Step (ii). By repeating Step (i) S times, we can use the empirical quantities of $\{\widetilde{\text{CD}}(I)^s : s = 1, \ldots, S\}$ to approximate $E[\text{CD}(I)|\mathcal{M}, \mathbf{Z}]$, $\text{Std}[\text{CD}(I)|\mathcal{M}, \mathbf{Z}]$, $Q_{CD(I)}(0.5|\mathcal{M}, \mathbf{Z})$, and $\text{Mstd}[\text{CD}(I)|\mathcal{M}, \mathbf{Z}]$. Subsequently, we can approximate $\text{CSCD}_1(I, \mathbf{Z})$ and $\text{CSCD}_2(I, \mathbf{Z})$ and determine their magnitude based on $\widetilde{\text{CD}}(I)^s$.
- Step (iii). We calculate two probabilities $P_A(I, \mathbf{Z}) = \sum_{s=1}^{S} \mathbf{1}(\widetilde{\text{CSCD}}_1(I, \mathbf{Z})^s \leq \widetilde{\text{CSCD}}_1(I, \mathbf{Z}))/S$ and $P_B(I, \mathbf{Z}) = \sum_{\tilde{I}} \sum_{s=1}^{S} \frac{\mathbf{1}(\widetilde{\text{CSCD}}_1(\tilde{I}, \mathbf{Z})^s \leq \widetilde{\text{CSCD}}_1(I, \mathbf{Z}))}{S \times \#(\tilde{I})}$, where $\#(\tilde{I})$ is the total number of all possible sets and $\mathbf{1}(\cdot)$ is an indicator function of a set.

Simulations

We generated 100 datasets from a linear mixed model as follows.

- Each dataset contains $n = 12$ clusters.
- For each cluster, $b_i \sim N(0, \sigma_b^2)$ and then, given b_i, y_{ij} $(j = 1, \cdots, m_i; i = 1, \ldots, n = 12)$ were independently generated from $N(\mathbf{x}_{ij}^T \beta + b_i, \sigma_y^2)$.
- m_i were randomly drawn from $\{1, \ldots, 5\}$.
- The covariates $\mathbf{x}_{ij}$ were set as $(1, u_i, t_{ij})^T$, where $t_{ij} = \log(j)$ and $u_i \sim N(0, 1)$.
- $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_b, \sigma_y)^T = (1, 1, 1, 1, 1)^T$.
- Consider the detection of influential clusters.

Simulations

- Scenario 1: simulation results from 100 datasets without influential clusters directly simulated from a linear mixed model.
- The $x-$axis corresponds to the order of the sorted degree of perturbation for all clusters.
- Panels (a), (b), and (c) show the box plots of $\mathrm{CD}(I)$, $E[\mathrm{CD}(I)|\mathcal{M}, \mathbf{Z}]$, and $\mathrm{Std}[\mathrm{CD}(I)|\mathcal{M}, \mathbf{Z}]$ as a function of $\mathcal{P}(I|M)$;
- panels (d), (e), and (f) show the box plots of $\mathrm{CD}(I) - \widetilde{\mathrm{CD}}(I)$, $E[\mathrm{CD}(I)|\mathcal{M}, \mathbf{Z}] - \widehat{M}[\widetilde{\mathrm{CD}}(I)]$, and $\mathrm{Std}[\mathrm{CD}(I)|\mathcal{M}, \mathbf{Z}] - \widehat{\mathrm{Std}}[\widetilde{\mathrm{CD}}(I)]$ as a function of $\mathcal{P}(I|M)$.

Simulations

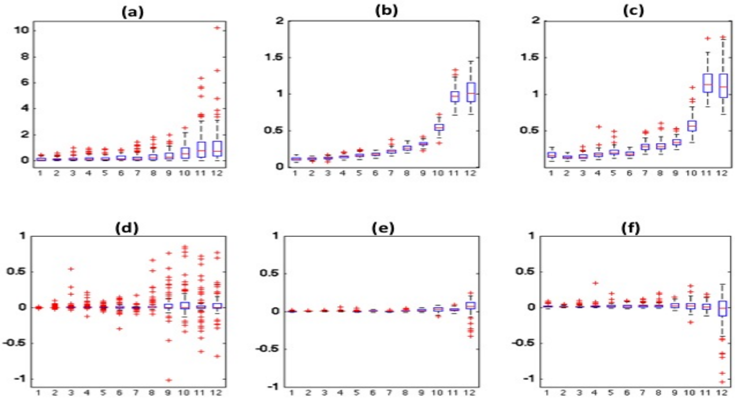

Figure: Scenario 1

Simulations

- Scenario 2: Simulation results from 100 datasets with two influential clusters simulated from a linear mixed model.
- We reset $(m_1, b_1) = (1, 4)$ and $(m_n, b_n) = (5, 3)$ to generate $y_{i,j}$ for $i = 1, n$ and all j according to the same linear mixed model.
- The $x-$axis corresponds to the order of the sorted degree of perturbation for all clusters.
- Panels (a), (b), and (c) show the box plots of $\text{CD}(I)$, $E[\text{CD}(I)|\mathcal{M}, \mathbf{Z}]$, and $\text{Std}[\text{CD}(I)|\mathcal{M}, \mathbf{Z}]$ as a function of $\mathcal{P}(I|M)$;
- panels (d), (e), and (f) show the box plots of $\text{CD}(I) - \widetilde{\text{CD}}(I)$, $E[\text{CD}(I)|\mathcal{M}, \mathbf{Z}] - \widehat{M}[\widetilde{\text{CD}}(I)]$, and $\text{Std}[\text{CD}(I)|\mathcal{M}, \mathbf{Z}] - \widehat{\text{Std}}[\widetilde{\text{CD}}(I)]$ as a function of $\mathcal{P}(I|M)$.

Simulations

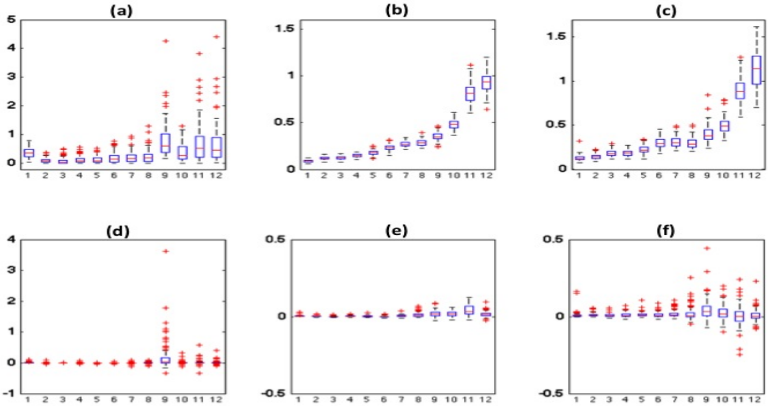

Figure: Scenario 2

Simulations

- Simulation results from 100 datasets simulated from a linear mixed model in the two scenarios.
- The first row corresponds to the first scenario, in which $m_{12} = 1$ and b_{12} varies from 0.6 to 6.0.
- The second row corresponds to the second scenario, in which $m_{12} = 10$ and b_{12} varies from 0.6 to 6.0.
- Panels (a) and (e) show the box plots of Cook's distances as a function of b_{12};
- panels (b) and (f) show the box plots of $\mathrm{CSCD}_1(I, \mathbf{Z})$ as a function of b_{12};
- panels (c) and (g) show the box plots of $P_B(I, \mathbf{Z})$ as a function of b_{12};
- panels (d) and (h) show the mean curve of $P_B(I, \mathbf{Z})$ based on $\mathrm{CSCD}_1(I, \mathbf{Z})$ (red line) and the mean curve of $P_C(I, \mathbf{Z})$ based on $\mathrm{CD}(I)$ (green line) as functions of b_{12}.

Simulations

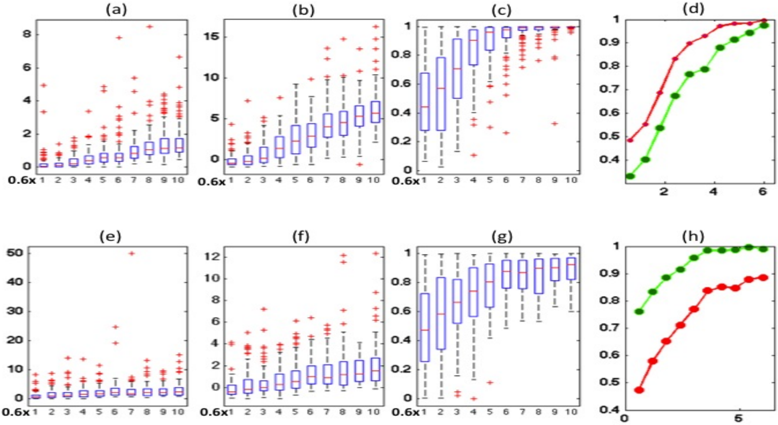

Figure: Scenario 3

Yale Infant Growth Data

- Study whether cocaine exposure during pregnancy may lead to the maltreatment of infants after birth.
- A total of 298 children were recruited from cocaine exposed group and unexposed group.
- $\sum_{i=1}^{n} m_i = 3176$, whereas m_i varies from 2 to 30.
- $y_{i,j} = \mathbf{x}_{i,j}^T\boldsymbol{\beta} + \epsilon_{i,j}$, where $y_{i,j}$ is the weight (in kilograms) of the j-th visit from the i-th subject, $\mathbf{x}_{i,j} = (1, d_{i,j}, (d_{i,j} - 120)^+, (d_{i,j} - 200)^+, (g_i - 28)^+, d_{i,j}(g_i - 28)^+, (d_{i,j} - 60)^+(g_i - 28)^+, (d_{i,j} - 490)^+(g_i - 28)^+, s_i d_{i,j}, s_i(d_{i,j} - 120)^+)^T$, in which $d_{i,j}$ and g_i (days) are the age of visit and gestational age, respectively, and s_i is the indicator for gender.
- $\boldsymbol{\epsilon}_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,m_i})^T \sim N_{m_i}(\mathbf{0}, R_i(\boldsymbol{\alpha}))$.
- M_1: $R_i(\boldsymbol{\alpha}) = \alpha_0 \mathbf{I}_{m_i} + \alpha_1 \mathbf{1}_{m_i}^{\otimes 2}$.
- M_2: $V(d) = \exp(\alpha_0 + \alpha_1 d + \alpha_2 d^2 + \alpha_3 d^3)$ and $\rho(l) = \alpha_4 + \alpha_5 l$, where l is the lag between two visits.

Yale Infant Growth Data

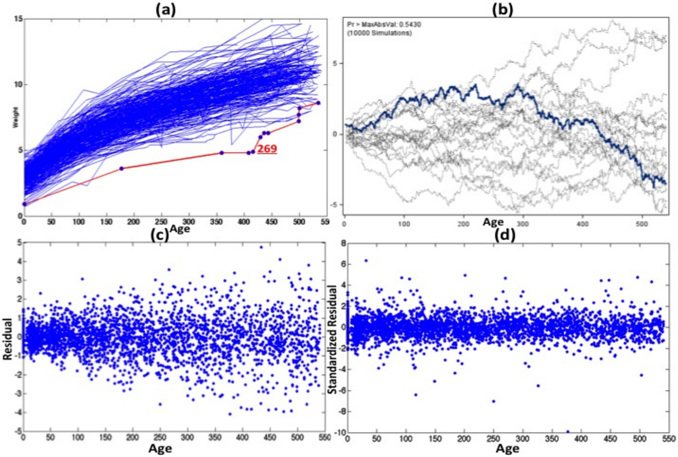

Figure: Panel (a): the line plot of infant weight against age; panel (b): the cumulative residual curve versus age; and panels (c) and (d): age versus raw residual and age versus studentized residual for cluster deletion.

Yale Growth Data

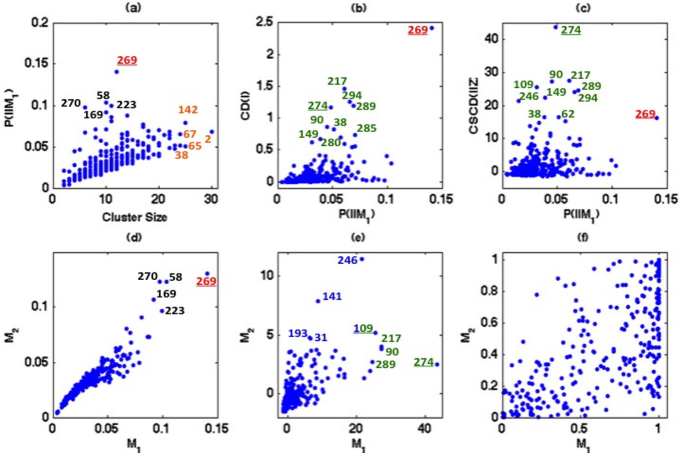

Figure: Panel (a): m_i versus $\mathcal{P}(I|\mathcal{M}_1)$; panel (b) shows $\mathcal{P}(I|\mathcal{M}_1)$ versus $\text{CD}(I)$; panel (c) shows $\mathcal{P}(I|\mathcal{M}_1)$ versus $\text{CSCD}_1(I, \mathbf{Z})$; and panels (d), (e), and (f): $\mathcal{P}(I|\mathcal{M})$, $\text{CSCD}_1(I, \mathbf{Z})$, and $P_B(I, \mathbf{Z})$ for models $\mathcal{M}_1$ and $\mathcal{M}_2$.

Yale Growth Data

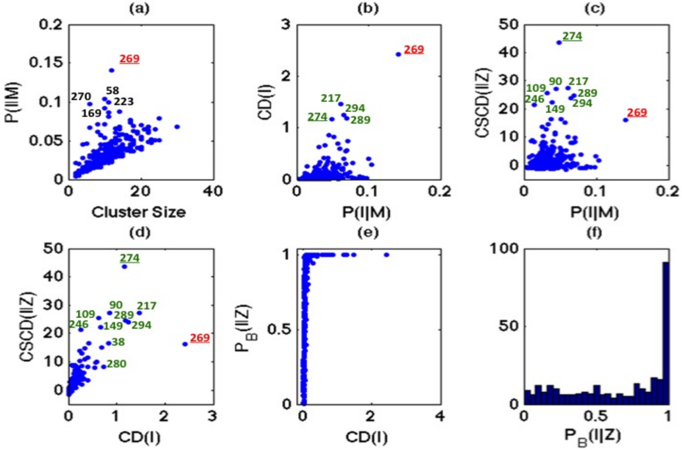

Figure: $\mathcal{M}_1$: Panel (a): $\mathcal{P}(I|M)$ versus m_i; panels (b), (c), (d), and (e): CD(I) versus $\mathcal{P}(I|M)$, CSCD($I|\mathbf{Z}$) versus $\mathcal{P}(I|M)$, CSCD($I|\mathbf{Z}$) versus CD(I), and $P_B(I|\mathbf{Z})$ versus CD(I); panel (f): the histogram of $P_B(I|\mathbf{Z})$.

Yale Growth Data

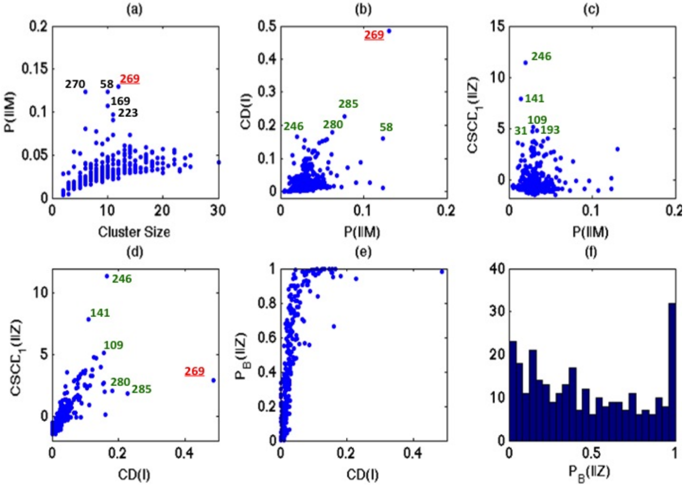

Figure: $\mathcal{M}_2$: Panel (a): $\mathcal{P}(I|M)$ versus m_i; panels (b), (c), (d), and (e): CD(I) versus $\mathcal{P}(I|M)$, CSCD($I|\mathbf{Z}$) versus $\mathcal{P}(I|M)$, CSCD($I|\mathbf{Z}$) versus CD(I), and

References

- Zhu, HT., Ibrahim JG, Cho HS. Perturbation and scaled Cook's distance. *Annals of Statistics*, 40, 785-811, 2012.
- Zhu, HT, Ibrahim JG., Cho HS, and Tang, N.S. Bayesian case-deletion measures for statistical models with missing data, *Journal of Computational and Graphical Statistics*, in press, 2012.
- Zhu HT, Ibrahim, JG., Shi, X.Y. Diagnostic measures for generalized linear models with missing covariates. *Scandivian Journal of Statistics*, 36, 686-712, 2009.
- Cho HS, Ibrahim JG, Sinha and Zhu HT. Bayesian Case Influence Diagnostics for Survival Models, *Biometrics*, 65, 116-124, 2009.
- Zhu HT, Tang NS, Ibrahim JG, and Zhang HP. Diagnostic measures for empirical likelihood of general estimating equations. *Biometrika*, 95, 489-507, 2008.
- Zhu HT, Lee SY, Wei BC, and Zhou JL. Case-deletion measures for models with incomplete data. *Biometrika*, 88:727-737, 2001.

Motivation

- Bayesian inference about a parameter $\boldsymbol{\theta}$ is typically based on calculating and summarizing the posterior distribution

$$p(\boldsymbol{\theta}|D_{obs}) = \frac{p(\boldsymbol{\theta})p(D_{obs}|\boldsymbol{\theta})}{\int p(\boldsymbol{\theta})p(D_{obs}|\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{3}$$

- It is well known that posterior quantities, such as the Bayes factor, posterior mean, etc... for a given dataset may be sensitive to any perturbation to the three key elements of a Bayesian analysis: D_{obs}, $p(\boldsymbol{\theta})$ and $p(D_{obs}|\boldsymbol{\theta})$.
- In the Bayesian literature, various methods for sensitivity analysis have been developed to perturb each of these three key elements and to assess the influence of various perturbations on the posterior distribution and its associated posterior quantities.

Motivation

- There are two major formal sensitivity techniques including the global and local robustness approaches (Berger, 1994).
- The key idea of the global robustness approach is to compute the range of posterior quantities as the perturbation to each of the three key elements change in a certain set of distributions and then determine the "extremal" ones (Berger, 1990).
- The conditional predictive ordinate (CPO) and the Kullback-Leibler divergence are two global influence measures for assessing individual observations.
- The Bayes factor can be regarded as a global sensitivity method.
- All these global sensitivity methods are generally computationally intensive for high-dimensional parameters.

Motivation

- The local robustness approach primarily computes the derivatives of the posterior quantities with respect to a small perturbation to $p(\boldsymbol{\theta})$ or $p(D_{obs}|\boldsymbol{\theta})$.
- In the frequentist literature, Cook's (1986) seminal local influence approach is particularly useful for perturbing $p(D_{obs}|\boldsymbol{\theta})$ in order to detect influential observations and assessing model misspecification.
- In the Bayesian literature, an analogue of Cook's (1986) approach has been developed (Gustafson, 1996; Gustafson and Wasserman, 1995; McCulloch, 1989; Berger, 1994; Berger, Insua, and Ruggeri, 2000).

Motivation

- Very little has been done on developing a general Bayesian influence approach for *simultaneously* perturbing D_{obs}, $p(\boldsymbol{\theta})$ and $p(D_{obs}|\boldsymbol{\theta})$, assessing their effects, and examining their applications in several settings, such as settings with missing data.
- Clarke and Gustafson (1998) is the sole paper on simultaneously perturbing $(D_{obs}, p(\boldsymbol{\theta}), p(D_{obs}|\boldsymbol{\theta}))$ in the context of independent and identically distributed data.

Motivation

We address three important issues related to the Bayesian influence approach:

- the development of a perturbation model that unifies various perturbation schemes for individually or simultaneously perturbing $(D_{obs}, p(\boldsymbol{\theta}), p(D_{obs}|\boldsymbol{\theta}))$;
- the development of a Bayesian perturbation manifold to characterize the intrinsic structure of the perturbation model;
- the development of local influence measures for selecting the most influential perturbation based on various objective functions and their statistical properties;
- the development of global influence measures for carrying out sensitivity analysis in missing data problem.

Perturbation Model

- Bayesian analysis of models with missing data:

$$p(\boldsymbol{\theta}|D_{obs}) \propto p(D_{obs};\boldsymbol{\theta})p(\boldsymbol{\theta}) \propto \int p(D_{com};\boldsymbol{\theta})p(\boldsymbol{\theta})\ d\Lambda(D_{mis}),$$

  where $\Lambda(\cdot)$ is an $\sigma-$finite measure, D_{obs} and D_{mis} are the observed data and the missing data, respectively, and $D_{com} = (D_{mis}, D_{obs})$ denotes the complete data.
- We develop a perturbation model to characterize various perturbation schemes to D_{com}, $p(D_{com};\boldsymbol{\theta})$ and $p(\boldsymbol{\theta})$.
- We embed all perturbed models in $\mathcal{P}_2$ and fix the initial model as the 'central point' of $\mathcal{P}_2$, where

$$\mathcal{P}_2 = \{p(\mathbf{s}) : R^{d_0} \to [0,\infty) | \int p(\mathbf{s})d\Lambda(\mathbf{s}) = 1\}, \tag{4}$$

  and d_0 is the dimension of $(D_{com}, \boldsymbol{\theta})$.

Perturbation Model

- We propose a perturbation model to the prior defined by

$$\mathcal{P}(\boldsymbol{\omega}_P) = \{p(\boldsymbol{\theta}, \boldsymbol{\omega}_P(\boldsymbol{\theta}))p(D_{com}; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta, \boldsymbol{\omega}_P(\cdot) \in \mathcal{L}_P\} \subset \mathcal{P}_2, \quad (5)$$

  where $\boldsymbol{\omega}_P(\cdot)$ is a $d_1 \times 1$ vector of real functions and $\mathcal{L}_P$ is a set of functions which map from Θ to R^{d_1}.
- The $\mathcal{L}_P$ may be infinite dimensional and $\boldsymbol{\omega}_P^0(\boldsymbol{\theta})$ in $\mathcal{L}_P$ represents no perturbation to the prior, that is $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}, \boldsymbol{\omega}_P^0(\boldsymbol{\theta}))$.
- This perturbation model includes *the additive $\epsilon-$contamination class, the geometric contamination class, and the parametric family* as special cases (Berger, 1990, 1994; Gustafson and Wasserman, 1995; Moreno, 2000).

Perturbation Model

- For example, consider

$$\boldsymbol{\beta} \sim N(\mu_0 + \boldsymbol{\omega}_{P,1}, \omega_{P,2}\Sigma_0),$$

  where $\boldsymbol{\omega}_{P,1} \in R^p$ and $\omega_{P,2} \geq 0$ is a positive scalar.
- Thus, $\boldsymbol{\omega}_P(\boldsymbol{\theta}) = (\boldsymbol{\omega}'_{P,1}, \omega_{P,2})' \in R^p \times [0, \infty)$ is independent of $\boldsymbol{\theta}$ and $d_1 = p + 1$.
- In this case, $\boldsymbol{\omega}^0_P(\boldsymbol{\theta}) = (\mathbf{0}'_p, 1)'$ represents no perturbation.

Perturbation Model

- The additive $\epsilon-$contamination class is given by

$$p(\boldsymbol{\theta}; \boldsymbol{\omega}_P(\boldsymbol{\theta})) = p(\boldsymbol{\theta}) + \epsilon[g(\boldsymbol{\theta}) - p(\boldsymbol{\theta})],$$

  where $\epsilon \in [0, 1]$ and $g(\boldsymbol{\theta})$ belongs to a class of contaminating distributions, denoted by $\mathcal{G}$ (Berger, 1994).

- We set

$$\boldsymbol{\omega}_P(\boldsymbol{\theta}) = \epsilon[g(\boldsymbol{\theta}) - p(\boldsymbol{\theta})],$$

  where $(\epsilon, g(\cdot))$ varies in $[0, 1] \times \mathcal{G}$. Thus, $\boldsymbol{\omega}_P^0(\boldsymbol{\theta}) = 0$.

- Similarly, the perturbation model (5) includes other perturbation schemes to the prior, such as the general $\epsilon-$contamination class and a general geometric contamination class (Perez, Martin, and Rufo, 2006)

Data and Sampling Distribution

- We propose a perturbation model to the complete-data sampling distribution

$$\mathcal{P}(\boldsymbol{\omega}_S) = \{p(\boldsymbol{\theta})p(D_{com}, \boldsymbol{\omega}_S(D_{com}, \boldsymbol{\theta}); \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta, \boldsymbol{\omega}_S(\cdot) \in \mathcal{L}_S\} \subset \mathcal{P}_2, \quad (6)$$

where $\boldsymbol{\omega}_S(D_{com}, \boldsymbol{\theta})$ is a function of D_{com} and $\boldsymbol{\theta}$ which belongs to the function space $\mathcal{L}_S$, for which $\boldsymbol{\omega}_S^0(\cdot) \in \mathcal{L}_S$ represents no perturbation.

- $\mathcal{P}(\boldsymbol{\omega}_S)$ automatically determines a perturbation model to the observed-data sampling distribution

$$\mathcal{P}_{obs}(\boldsymbol{\omega}_S) = \left\{ \int p(D_{com}, \boldsymbol{\omega}_S(D_{com}, \boldsymbol{\theta}); \boldsymbol{\theta}) d\Lambda(D_{mis}) : p(D_{com}, \boldsymbol{\omega}_S(D_{com}, \boldsymbol{\theta}); \boldsymbol{\theta}) \in \mathcal{P}(\boldsymbol{\omega}_S) \right\}.$$

Simultaneous Perturbation

- We propose a perturbation model to simultaneously perturb the data, the prior, and the sampling distribution

$$\mathcal{P}(\boldsymbol{\omega}) = \{p(\boldsymbol{\theta}, \boldsymbol{\omega}_P(\boldsymbol{\theta}))p(D_{com}, \boldsymbol{\omega}_S(D_{com}, \boldsymbol{\theta}); \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta, \boldsymbol{\omega}(\cdot) \in \mathcal{L}\} \subset \mathcal{P}_2, \quad (7)$$

where

$$\boldsymbol{\omega} = \boldsymbol{\omega}(D_{com}, \boldsymbol{\theta}) = (\boldsymbol{\omega}_P(\boldsymbol{\theta}), \boldsymbol{\omega}_S(D_{com}, \boldsymbol{\theta})) \in \Omega = \mathcal{L}_P \times \mathcal{L}_S$$

and

$$\boldsymbol{\omega}^0(D_{com}, \boldsymbol{\theta}) = (\boldsymbol{\omega}_P^0(\boldsymbol{\theta}), \boldsymbol{\omega}_S^0(D_{com}, \boldsymbol{\theta}))$$

is the 'central point' of Ω representing no perturbation.

- $\mathcal{P}(\boldsymbol{\omega}_P, \boldsymbol{\omega}_S^0) = \mathcal{P}_P(\boldsymbol{\omega}_P)$ and $\mathcal{P}(\boldsymbol{\omega}_P^0, \boldsymbol{\omega}_S) = \mathcal{P}_S(\boldsymbol{\omega}_P)$ represent the individual perturbations to the prior and the sampling distribution, respectively.

Simultaneous Perturbation

- Based on the perturbation model (7), we can measure the amount of perturbation, the extent to which each component of a perturbation model contributes to, and the degree of orthogonality for the components of the perturbation model.
- Such a quantification is very useful for rigorously assessing the relative influence of each component in a Bayesian analysis, which can reveal any discrepancy among data, the prior, or the sampling model.
- For instance, a data-prior discrepancy can arise when either an estimate of the parameter is in a low probability region of the prior or the prior leads to an improper posterior distribution.
- Because the components of the perturbation model may not be orthogonal to each other, special care should be taken when we interpret local influence measures from such a perturbation.

Bayesian Perturbation Manifold

- We develop a Bayesian perturbation manifold (BPM) to measure each perturbation $\boldsymbol{\omega}$ in the perturbation model and apply this methodology to a wide variety of statistical models, allowing for incomplete-data.
- The perturbation model $\mathcal{M} = \{p(D_{com}, \boldsymbol{\theta}; \boldsymbol{\omega}) : \boldsymbol{\omega} \in \Omega\}$ has a natural geometrical structure. Since Ω can be either a finite dimensional set or an infinite dimensional set, we need to develop a manifold for infinite dimensional space, which includes the finite dimensional manifold as a submanifold.
- For instance, Ω for the $\epsilon-$contamination class and the linear perturbation class are infinite dimensional, whereas Ω for the parametric family are finite dimensional.

Bayesian Perturbation Manifold

- When Ω is an infinite dimensional set,

$$\mathcal{M} = \{p_c(\boldsymbol{\omega}) = p(D_{com}, \boldsymbol{\theta}; \boldsymbol{\omega}) : \boldsymbol{\omega} \in \Omega\} \subset \mathcal{P}_2 \tag{8}$$

  is an infinite dimensional manifold (Lang, 1995; Friedrich, 1991; Zhang, 2007).

- Assume that

$$C(t) : p_c(\boldsymbol{\omega}(t)) = p(D_{com}, \boldsymbol{\theta}; \boldsymbol{\omega}(t))$$

  is a differentiable function mapping from $t \subset I \in R$ to the manifold $\mathcal{M}$ with

$$p_c(\boldsymbol{\omega}(0)) = p(D_{com}, \boldsymbol{\theta}; \boldsymbol{\omega}),$$

  where I is an open interval covering 0.

Bayesian Perturbation Manifold

- Let

$$\dot{p}_c(\omega(t)) = dp_c(\omega(t))/dt$$

and let $P(\omega)$ be the probability measure determined by $p_c(\omega)$ such that

$$\frac{dP(\omega)}{d\Lambda(D_{com}, \boldsymbol{\theta})} = p_c(\omega).$$

Bayesian Perturbation Manifold

- At each ω, there is a tangent space $T_\omega\mathcal{M}$ of $\mathcal{M}$ defined by

$$T_\omega\mathcal{M} = \{\mathbf{v}(\omega) = \dot{p}_c(\omega(0)) : \int \mathbf{v}(\omega) d\Lambda(D_{com}, \boldsymbol{\theta}) = 0 \text{ and } \mathbf{v}(\omega)/p_c(\omega) \in L^2(P(\omega)) < \infty\}, \quad (9)$$

  where $L^2(P(\omega)) = \{g : \int g^2 dP(\omega) < \infty\}$ is a Hilbert space.
- The inner product of $\mathbf{v}_1(\omega)$ and $\mathbf{v}_2(\omega)$ in $T_\omega\mathcal{M}$ is defined as

$$g(\mathbf{v}_1, \mathbf{v}_2)(\omega) = \int \frac{\mathbf{v}_1(\omega)}{p_c(\omega)} \frac{\mathbf{v}_2(\omega)}{p_c(\omega)} p_c(\omega) d\Lambda(D_{com}, \boldsymbol{\theta}). \quad (10)$$

Bayesian Perturbation Manifold

- The length of the curve C from t_1 to t_2 is given by

$$S_C(\omega(t_1), \omega(t_2)) = \int_{t_1}^{t_2} \sqrt{g(\dot{p}_c(\omega(t)), \dot{p}_c(\omega(t)))}dt. \qquad (11)$$

  The tangent manifold $\mathcal{TM} = \cup_{\omega \in \mathcal{M}} T_\omega \mathcal{M}$ is the disjoint union of the tangent spaces for all points on $\mathcal{M}$.
- To define the notion of 'straight line' on $\mathcal{M}$, we need to introduce the concepts of Levi-Civita connection and geodesic.
- Let $\mathbf{u}(\omega) = \mathbf{u}(p_c(\omega))$ and $\mathbf{v}(\omega) = \mathbf{v}(p_c(\omega))$ be two smooth vector fields defined from $\mathcal{M}$ to $\mathcal{TM}$.

Bayesian Perturbation Manifold

- We define the directional derivative $d\mathbf{u}[\mathbf{v}]$ of a vector field $\mathbf{u}$ in the direction of $\mathbf{v}(\omega) \in T_{\omega(0)}\mathcal{M}$ at $\omega(0) = \omega$ as

$$d\mathbf{u}[\mathbf{v}](\omega) = \lim_{t\to 0} t^{-1}[\mathbf{u}(\omega(t)) - \mathbf{u}(\omega(0))]. \quad (12)$$

- The covariant derivative for Levi-Civita connection $\nabla_{\mathbf{v}}\mathbf{u}$ is given by

$$\nabla_{\mathbf{v}}\mathbf{u}(\omega) = d\mathbf{u}[\mathbf{v}](\omega) - 0.5\{\mathbf{u}(\omega)\mathbf{v}(\omega)[p_c(\omega)]^{-1} - \int \mathbf{u}(\omega)\mathbf{v}(\omega)[p_c(\omega)]^{-1} d\Lambda(D_{com}, \boldsymbol{\theta})\}.$$

Bayesian Perturbation Manifold

- A geodesic on the manifold $(\mathcal{M}, g(\cdot,\cdot))$ is a smooth map $\gamma(t)$ from (a,b) to $\mathcal{M}$ such that $\nabla_{\dot\gamma(t)}\dot\gamma(t) = 0$, where $\dot\gamma(t) = d\gamma(t)/dt$.
- For every $\mathbf{u} \in T_{\boldsymbol\omega}\mathcal{M}$, there is a unique geodesic $\gamma(t;\boldsymbol\omega,\mathbf{u}) : I \to \mathcal{M}$ passing through $\gamma(0;\boldsymbol\omega,\mathbf{u}) = \boldsymbol\omega$ with initial direction $\dot\gamma(0;\boldsymbol\omega,\mathbf{u}) = \mathbf{u}$, where

$$\dot\gamma(t;\boldsymbol\omega,\mathbf{u}) = d\gamma(t;\boldsymbol\omega,\mathbf{u})/dt$$

  and I is an open interval containing 0.
- The geodesic is a direct extension of the straight line

$$\boldsymbol\omega(t) = \boldsymbol\omega^0 + t\mathbf{h}$$

  in finite dimensional Euclidean space (Amari, 1990; Kass and Vos, 1997).

Bayesian Perturbation Manifold

- DEFINITION 1. A Bayesian perturbation manifold $(\mathcal{M}, g(\mathbf{u}, \mathbf{v}), \nabla_{\mathbf{v}}\mathbf{u})$ is the manifold $\mathcal{M}$ with an inner product $g(\mathbf{u}, \mathbf{v})$ and a covariant derivative for the Levi-Civita connection $\nabla_{\mathbf{v}}\mathbf{u}$.
- When $\Omega \subset R^m$, $\mathcal{M}_m = \{p_c(\boldsymbol{\omega}) = p(D_{com}, \boldsymbol{\theta}; \boldsymbol{\omega}) : \boldsymbol{\omega} \in \Omega\}$ is an $m-$dimensional submanifold of the infinite-dimensional manifold $\mathcal{M}$.
- The tangent vector field of $\mathcal{M}_m$ takes the form $\mathbf{u} = \partial_{\omega_j} p_c(\boldsymbol{\omega})$ and the tangent space $\mathcal{T}_{\boldsymbol{\omega}}\mathcal{M}_m$ is spanned by the m tangent vectors $\partial_{\omega_j} p_c(\boldsymbol{\omega})$.
- $$g_{jk}(\boldsymbol{\omega}) = \int [\partial_{\omega_j}\ell_c(\boldsymbol{\omega})][\partial_{\omega_k}\ell_c(\boldsymbol{\omega})]p_c(\boldsymbol{\omega})d\Lambda(D_{com}, \boldsymbol{\theta}),$$

  where $\ell_c(\boldsymbol{\omega}) = \log p_c(\boldsymbol{\omega})$.

Bayesian Perturbation Manifold

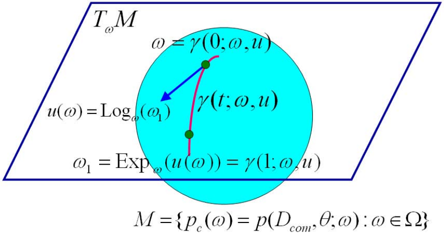

Figure: Graphical illustration of geodesic, exponential and logarithm maps. The map $\gamma(t; \boldsymbol{\omega}, \mathbf{u}) : I \to \mathcal{M}$ is a geodesic passing through $\gamma(0; \boldsymbol{\omega}, \mathbf{u}) = \boldsymbol{\omega}$ and $\gamma(1; \boldsymbol{\omega}, \mathbf{u}) = \boldsymbol{\omega}_1$ with initial direction $\dot{\gamma}(0; \boldsymbol{\omega}, \mathbf{u}) = \mathbf{u}$. The exponential and logarithm maps are, respectively, defined as $\boldsymbol{\omega}_1 = \mathrm{Exp}_{\boldsymbol{\omega}}(\mathbf{u})$ and $\mathbf{u}(\boldsymbol{\omega}) = \mathrm{Log}_{\boldsymbol{\omega}}(\boldsymbol{\omega}_1)$.

Bayesian Perturbation Manifold

- The Levi-Civita connection is defined as

$$\begin{aligned}\nabla_{\partial_{\omega_j} p_c(\boldsymbol{\omega})} \partial_{\omega_k} p_c(\boldsymbol{\omega}) &= \partial^2_{\omega_j \omega_k} p_c(\boldsymbol{\omega}) - 0.5\{[\partial_{\omega_j} p_c(\boldsymbol{\omega}) \partial_{\omega_k} p_c(\boldsymbol{\omega})] p_c(\boldsymbol{\omega})^{-1} \\ &- \int [\partial_{\omega_j} p_c(\boldsymbol{\omega}) \partial_{\omega_k} p_c(\boldsymbol{\omega})] p_c(\boldsymbol{\omega})^{-1} d\Lambda(D_{com}, \boldsymbol{\theta})\},\end{aligned}$$

and

$$\Gamma_{jkl}(\boldsymbol{\omega}) = g(\nabla_{\partial_{\omega_j} p_c(\boldsymbol{\omega})} \partial_{\omega_k} p_c(\boldsymbol{\omega}), \partial_{\omega_l} p_c(\boldsymbol{\omega}))$$

is the Christoffel symbol for $\nabla_{\partial_{\omega_j} p_c(\boldsymbol{\omega})} \partial_{\omega_k} p_c(\boldsymbol{\omega})$.

Examples of Bayesian Perturbation Manifolds

BPM for the Prior

- For the parametric family perturbation to the prior,

$$g_{jk}(\boldsymbol{\omega}_P) = \int [\partial_{\omega_j} \ell(\boldsymbol{\theta}; \boldsymbol{\omega}_P) \partial_{\omega_k} \ell(\boldsymbol{\theta}; \boldsymbol{\omega}_P)] p(\boldsymbol{\theta}; \boldsymbol{\omega}_P) d\Lambda(\boldsymbol{\theta}), \quad (13)$$

where $\ell(\boldsymbol{\theta}; \boldsymbol{\omega}_P) = \log p(\boldsymbol{\theta}; \boldsymbol{\omega}_P)$.

- We consider a hierarchical structure for the prior,

$$p(\boldsymbol{\theta}) = p(\theta_1) p(\theta_2; \boldsymbol{\theta}_{[1]}) \cdots p(\theta_p; \boldsymbol{\theta}_{[p-1]})$$

and

$$p(\boldsymbol{\theta}; \boldsymbol{\omega}_P) = p(\theta_1; \omega_{P,1}) p(\theta_2; \theta_{[1]}, \omega_{P,2}) \cdots p(\theta_p; \theta_{[p-1]}, \omega_{P,p}), \quad (14)$$

where $\boldsymbol{\theta}_{[j]} = (\theta_1, \cdots, \theta_{j-1})$.

Bayesian Perturbation Manifold

- Different $\omega_{P,j}$ are orthogonal to each other, that is $g_{jk}(\boldsymbol{\omega}) = 0$ for all $j \neq k$.
- All geometric quantities (e.g., geodesic) of the BPM for the prior are independent of the sampling distribution.

Examples of Bayesian Perturbation Manifolds

BPM for the $\epsilon-$contamination class of priors

- This BPM is an infinite dimensional manifold. Recall that

$$\omega_P(\boldsymbol{\theta}) = \epsilon[g(\boldsymbol{\theta}) - p(\boldsymbol{\theta})].$$

- We substitute ϵ with t in $\omega_P(\boldsymbol{\theta})$, which yields

$$\omega_P(\boldsymbol{\theta}) = \omega_P(t, g(\boldsymbol{\theta})) = t[g(\boldsymbol{\theta}) - p(\boldsymbol{\theta})].$$

- Considering

$$\mathbf{v}_j(\omega_P^0) = \dot{\omega}_P(t, g_j(\boldsymbol{\theta})) = [g_j(\boldsymbol{\theta}) - p(\boldsymbol{\theta})]p(D_{com}; \boldsymbol{\theta})$$

  for $j = 1, 2$, we get

$$g(\mathbf{v}_1, \mathbf{v}_2)(\omega_P^0) = \int [g_1(\boldsymbol{\theta})/p(\boldsymbol{\theta}) - 1][g_2(\boldsymbol{\theta})/p(\boldsymbol{\theta}) - 1]p(\boldsymbol{\theta})d\Lambda(\boldsymbol{\theta}),$$

  which is independent of $p(D_{com}; \boldsymbol{\theta})$.

Examples of Bayesian Perturbation Manifolds

- $$g(\mathbf{v},\mathbf{v})(\omega_P^0) = \int [g(\boldsymbol{\theta})/p(\boldsymbol{\theta}) - 1]^2 p(\boldsymbol{\theta}) d\Lambda(\boldsymbol{\theta})$$

  reduces to the L^2 norm considered in Gustafson (1996a).
- The BPMs for all perturbations to the prior are independent of the specification of the sampling distribution.

Examples of Bayesian Perturbation Manifolds

BPM for the single-case perturbation scheme to the sampling distribution

- For the independent-type-incomplete-data model, the complete-data density for the single-case perturbation may be defined by

$$p(D_{com};\boldsymbol{\theta},\boldsymbol{\omega}_S)=\prod_{i=1}^{n}p(\mathbf{d}_{i,c};\boldsymbol{\theta},\omega_{S,i}),$$

- This BPM is a finite dimensional manifold with metric tensor

$$g_{jk}(\boldsymbol{\omega}_S)=\int\tilde{g}_{jk}(\boldsymbol{\theta};\boldsymbol{\omega}_S)p(\boldsymbol{\theta})d\Lambda(\boldsymbol{\theta})$$

for $j,k=1,\cdots,n$, where

$$\tilde{g}_{jk}(\boldsymbol{\theta};\boldsymbol{\omega}_S)=\delta_{jk}\int[\partial_{\omega_{S,j}}\log p(\mathbf{d}_{j,c},\omega_{S,j};\boldsymbol{\theta})]^{\otimes 2}p(\mathbf{d}_{j,c};\boldsymbol{\theta})d\Lambda(D_{com}).$$

- If $p(\boldsymbol{\theta})$ concentrates on $\hat{\boldsymbol{\theta}}_{mle}$, then we obtain the metric tensor $g_{jk}(\boldsymbol{\omega}_S)=\tilde{g}_{jk}(\hat{\boldsymbol{\theta}}_{mle};\boldsymbol{\omega}_S)$ defined in Zhu et al. (2007).

Bayesian Perturbation Manifold

- THEOREM 1. $g(\mathbf{v},\mathbf{v})(\omega^0) = g_S(\mathbf{v},\mathbf{v})(\omega^0) + g_P(\mathbf{v},\mathbf{v})(\omega^0)$, *in which*

$$g_S(\mathbf{v},\mathbf{v})(\omega^0) = \int [d_t \log p(D_{com};\boldsymbol{\theta},\boldsymbol{\omega}_S(t))]^2 p_c(\boldsymbol{\omega}) d\Lambda(D_{com},\boldsymbol{\theta}),$$

$$g_P(\mathbf{v},\mathbf{v})(\omega^0) = \int [d_t \log p(\boldsymbol{\theta};\boldsymbol{\omega}_P(t))]^2 p(\theta) d\Lambda(\boldsymbol{\theta}). \qquad (15)$$

- For simultaneous perturbations to the prior, the data and the sampling distribution, if the components in $\boldsymbol{\omega}_P$ and $\boldsymbol{\omega}_S$ are different, then Theorem 1 ensures that the BPMs for $\boldsymbol{\omega}_P$ and $\boldsymbol{\omega}_S$ are geometrically orthogonal to each other.

Global Influence Measures

We develop several global influence measures for quantifying the effects of perturbing the three key elements of a Bayesian analysis.

- Let $p_c(\omega^0)$ and $p_c(\omega)$ represent the unperturbed and perturbed complete-data distributions.
- Let $C(t) = p_c(\omega(t)) : [-\delta, \delta] \to \mathcal{M}$ be a smooth curve on $\mathcal{M}$ joining $p_c(\omega^0)$ and $p_c(\omega(s))$ such that $C(0) = p_c(\omega^0)$ and $C(1) = p_c(\omega)$, where $\delta > 1$.
- We consider a smooth function of interest

$$f(\omega) = f(p_c(\omega)) : \mathcal{M} \to R$$

for sensitivity analysis. Thus,

$$f(\omega(t)) : [-\delta, \delta] \to R$$

is a real function of t.

Global Influence Measures

- DEFINITION 2. The *global influence measure* for comparing $p(\boldsymbol{\theta}|D_{obs}, \boldsymbol{\omega}^0)$ and $p(\boldsymbol{\theta}|D_{obs}, \boldsymbol{\omega})$ along the smooth curve $C(t)$ is defined as

$$\mathrm{GI}_{f,C(t)}(\boldsymbol{\omega}^0, \boldsymbol{\omega}) = \frac{[f(\boldsymbol{\omega}) - f(\boldsymbol{\omega}^0)]^2}{S_C(\boldsymbol{\omega}^0, \boldsymbol{\omega})^2}. \quad (16)$$

- The $\mathrm{GI}_{f,C}(\boldsymbol{\omega}^0, \boldsymbol{\omega})$ can be interpreted as the ratio of the change of the objective function over the length of the curve $C(t)$ on the manifold $\mathcal{M}$.
- THEOREM 2. $\mathrm{GI}_{f,C(t)}(\boldsymbol{\omega}^0, \boldsymbol{\omega})$ is invariant with respect to any reparametrizations of the curve $C(t)$.

Global Influence Measures

- $\text{GI}_{f,C(t)}(\omega^0, \omega)$ depends on the particular path $C(t)$.
- DEFINITION 3. The intrinsic global influence measure for comparing $p(\theta|D_{obs}, \omega^0)$ and $p(\theta|D_{obs}, \omega)$ is defined as

$$\text{IGI}_f(\omega^0, \omega) = \frac{[f(\omega) - f(\omega^0)]^2}{d(\omega^0, \omega)^2}. \quad (17)$$

- The proposed $\text{IGI}_f(\omega^0, \omega)$ can be interpreted as the ratio of the change of the objective function over the minimal distance between $p_c(\omega^0)$ and $p_c(\omega)$ on $\mathcal{M}$.
- THEOREM 3. If $\mathcal{M}$ is a complete Riemannian manifold, then

$$\text{IGI}_f(\omega^0, \omega) = \max_{C(t) \in \mathcal{L}(\omega^0, \omega)} \text{GI}_{f,C(t)}(\omega^0, \omega).$$

Global Influence Measures

DEFINITION 4. The global influence measure for comparing $p(\boldsymbol{\theta}|D_{obs}, \boldsymbol{\omega}^0)$ to all $p(\boldsymbol{\theta}|D_{obs}, \boldsymbol{\omega})$ for $\boldsymbol{\omega} \in \Omega_1$ along the smooth curve family $\{C(t; \boldsymbol{\omega}) : \boldsymbol{\omega} \in \Omega_1\}$ is defined as

$$\mathrm{GI}_{f,C(t)}(\boldsymbol{\omega}^0, \Omega_1) = \sup_{\boldsymbol{\omega} \in \Omega_1} \mathrm{GI}_{f,C(t;\boldsymbol{\omega})}(\boldsymbol{\omega}^0, \boldsymbol{\omega}). \tag{18}$$

The intrinsic global influence measure for comparing $p(\boldsymbol{\theta}|D_{obs}, \boldsymbol{\omega}^0)$ to all $p(\boldsymbol{\theta}|D_{obs}, \boldsymbol{\omega})$ for $\boldsymbol{\omega} \in \Omega_1$ is defined as

$$\mathrm{IGI}_f(\boldsymbol{\omega}^0, \Omega_1) = \max_{\boldsymbol{\omega} \in \Omega_1} \mathrm{IGI}_f(\boldsymbol{\omega}^0, \boldsymbol{\omega}). \tag{19}$$

Local Influence Measures

- $$f(\omega(t)) = f(\omega(0)) + \dot{f}(\omega(0))t + 0.5\ddot{f}(\omega(0))t^2 + o(t^2).$$
- We need to distinguish two cases: $\dot{f}(\omega(0)) \neq 0$ for some smooth curves $\omega(t)$ and $\dot{f}(\omega(0)) = 0$ for all smooth curves $\omega(t)$. If $\dot{f}(\omega(0)) = 0$ for all smooth curves $\omega(t)$, then we have to consider the second order term $\ddot{f}(\omega(0))$ in order to characterize the local behavior of $f(\omega(t))$.
- Definition 5. The first-order local influence measure is defined as

$$\mathrm{FI}_f[\mathbf{v}](\omega(0)) = \lim_{t\to 0} \mathrm{GI}_{f,C(t)}(\omega(0), \omega(t)) = \frac{[df[\mathbf{v}](\omega(0))]^2}{g(\mathbf{v},\mathbf{v})(\omega(0))}. \quad (20)$$

Local Influence Measures

- $\mathrm{FI}_f[\mathbf{v}](\omega(0))$ is invariant with respect to any reparametrizations of the curve $\omega(t)$.
- For any finite-dimensional manifold, we have

$$\mathrm{FI}_f[\mathbf{v}](\omega(0)) = \frac{[\mathbf{v}_h^T \partial_\omega f(\omega(0))]^2}{\mathbf{v}_h^T G(\omega(0))\mathbf{v}_h}, \qquad (21)$$

where $\mathbf{v}_h = (v_1, \cdots, v_p)$ equals $d_t\omega(t)$ evaluated at $t = 0$. Moreover, if ϕ is a diffeomorphism of ω, then $\mathrm{FI}_f[\mathbf{v}](\omega(0))$ is invariant with respect to any reparametrization corresponding to ϕ.

Local Influence Measures

- For a finite-dimensional manifold, we use the direction vector

$$\mathbf{v}_{\max} = [G(\boldsymbol{\omega}(0))]^{-1/2}\partial_{\boldsymbol{\omega}} f(\boldsymbol{\omega}(0))$$

instead of $\text{grad}(f)(\boldsymbol{\omega}(0))$ to identify influential directions, since

$$[G(\boldsymbol{\omega}(0))]^{-1/2}\partial_{\boldsymbol{\omega}} p_c(\boldsymbol{\omega}(0))$$

forms an orthonormal basis at $\boldsymbol{\omega}(0)$.

Local Influence Measures

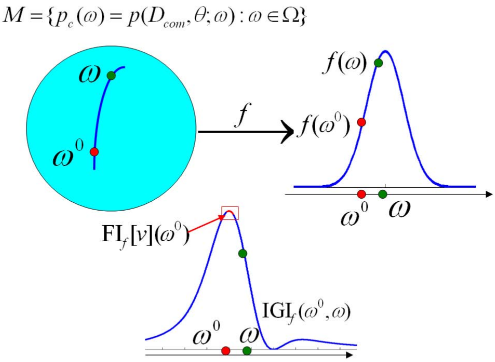

Figure: Graphical illustration of objective function $f(\boldsymbol{\omega}(t))$, intrinsic global influence measure, $\mathrm{IGI}_f(\boldsymbol{\omega}^0, \boldsymbol{\omega})$, and local influence measure $\mathrm{FI}_f[\mathbf{v}](\boldsymbol{\omega}^0)$: $\mathrm{FI}_f[\mathbf{v}](\boldsymbol{\omega}^0)$ quantifies the local behavior of $f(\boldsymbol{\omega}(t))$ near $\boldsymbol{\omega}^0$ and $\mathrm{IGI}_f(\boldsymbol{\omega}^0, \boldsymbol{\omega})$ quantifies the relative change of $f(\boldsymbol{\omega})$ relative to the shortest distance between $\boldsymbol{\omega}^0$ and $\boldsymbol{\omega}$.

Local Influence Measures

- We only consider the geodesic $p_c(\omega(t)) = \text{Exp}_{p_c(\boldsymbol{\omega}(0))}(t\mathbf{v})$ that satisfies $p_c(\omega(0)) = p_c(\omega^0)$ and $d_t p_c(\omega(0)) = \mathbf{v} \in T_{\boldsymbol{\omega}(0)}\mathcal{M}$.
- We obtain a covariant version of Taylor's theorem as follows:

$$f(\text{Exp}_{\boldsymbol{\omega}(0)}(t\mathbf{v})) = f(\omega^0) + tdf[\mathbf{v}](\omega(0)) + 0.5t^2\text{Hess}(f)(\mathbf{v},\mathbf{v})(\omega(0)) + o(t^2),$$

where $\text{Hess}(f)(\mathbf{v},\mathbf{v})(\omega(0)) = \ddot{f}(\text{Exp}_{\boldsymbol{\omega}(0)}(t\mathbf{v}))|_{t=0}$ is a covariant (or Riemmanian) Hessian.

- Geometrically, $\text{Hess}(f)(\mathbf{u},\mathbf{v})(\omega(0))$ is a tensor of type (0,2) and defined as

$$\text{Hess}(f)(\mathbf{u},\mathbf{v})(\omega(0)) = d(df[\mathbf{v}])[\mathbf{u}](\omega(0)) - df[\nabla_{\mathbf{u}}\mathbf{v}](\omega(0)). \quad (22)$$

Local Influence Measures

- The Hessian $\text{Hess}(f)(\mathbf{v},\mathbf{v})(\omega(0))$ is invariant with respect to any functions satisfying $\omega(0)=\omega^0$ and $d_t p_c(\omega(0))=\mathbf{v}\in T_{\boldsymbol{\omega}(0)}\mathcal{M}$.
- DEFINITION 6. The second-order influence measure (SI) in the direction $\mathbf{v}\in T_{\boldsymbol{\omega}(0)}\mathcal{M}$ is defined as

$$\text{SI}_f[\mathbf{v}](\omega(0))=\frac{\text{Hess}(f)(\mathbf{v},\mathbf{v})(\omega(0))}{g(\mathbf{v},\mathbf{v})(\omega(0))}. \tag{23}$$

- $\text{SI}_f[k\mathbf{v}](\omega(0))=\text{SI}_f[\mathbf{v}](\omega(0))$ for any $k\neq 0$.

Local Influence Measures

- For a finite dimensional manifold, $\text{Hess}(f)(\mathbf{v}, \mathbf{v})(\omega(0))$ reduces to $\mathbf{v}_h^T H_f(\omega(0))\mathbf{v}_h$, where the $(j, k)^{th}$ element of $H_f(\omega)$ is given by

$$[H_f(\omega)]_{(j,k)} = \partial^2_{\omega_j \omega_k} f(\omega) - \sum_{s,r} g^{sr}(\omega)\Gamma_{jks}(\omega)\partial_{\omega_r} f(\omega). \qquad (24)$$

- $\text{SI}_f[\mathbf{v}](\omega(0))$ is invariant with respect to any reparametrization corresponding to ϕ at $\omega(0)$.

Theoretical Examples

Bayes Factor

- $f(\omega) = B(\omega)$ and $\omega(t)$ is a smooth curve on $\mathcal{M}$ with $\omega(0) = \omega^0$ and $d_t p_c(\omega(t))|_{t=0} = \mathbf{v} \in T_{\omega(0)}\mathcal{M}$.
-
$$B(\omega) = \log p(D_{obs}; \omega) - \log p(D_{obs}; \omega^0)$$
is a continuous map from $\mathcal{M}$ to R.
- We consider the simultaneous perturbation to both the prior and the sampling distribution and, therefore we have

$$\text{FI}_B[\mathbf{v}](\omega(0)) = \frac{E[d_t \log p(D_{com}, \boldsymbol{\theta}; \omega(t))|D_{obs}]^2}{g_P(\mathbf{v}, \mathbf{v}) + g_S(\mathbf{v}, \mathbf{v})}. \qquad (25)$$

Theoretical Examples

- For $p(\boldsymbol{\theta}; t) = p(\boldsymbol{\theta}) + t[g(\boldsymbol{\theta}) - p(\boldsymbol{\theta})]$, we have

$$\mathrm{FI}_B[\mathbf{v}](\boldsymbol{\omega}(0))] = \frac{\mathrm{E}[g(\boldsymbol{\theta})/p(\boldsymbol{\theta})|D_{obs}]^2}{\mathrm{Var}_P(g(\boldsymbol{\theta})/p(\boldsymbol{\theta}))} = \frac{[p_g(D_{obs})/p(D_{obs})]^2}{\mathrm{Var}_P(g(\boldsymbol{\theta})/p(\boldsymbol{\theta}))}, \quad (26)$$

  where

$$p(D_{obs}) = \int p(D_{com}; \boldsymbol{\theta})p(\boldsymbol{\theta})d\Lambda(D_{mis}, \boldsymbol{\theta})$$

  and

$$p_g(D_{obs}) = \int p(D_{com}; \boldsymbol{\theta})g(\boldsymbol{\theta})d\Lambda(D_{mis}, \boldsymbol{\theta}).$$

- $\mathrm{FI}_B[\mathbf{v}](\boldsymbol{\omega}(0))]$ is the square of the normalized Bayes factor of $g(\boldsymbol{\theta})$ against $p(\boldsymbol{\theta})$.

Theoretical Examples

Bayes Factor

- For a perturbation scheme to the sampling distribution,

$$\dot{f}(\omega(0)) = E[d_t \ell_c(\hat{\theta}, \omega(0)) | D_{obs}] \approx d_t \ell_o(\hat{\theta}, \omega(0))$$

and

$$\mathrm{FI}_B[\mathbf{v}](\omega(0))] = \frac{E[d_t \ell_c(\hat{\theta}, \omega(0)) | D_{obs}]^2}{g_S(\mathbf{v}, \mathbf{v})} \approx \frac{[d_t \ell_o(\hat{\theta}, \omega(0))]^2}{g_S(\mathbf{v}, \mathbf{v})},$$

where

$$\ell_o(\hat{\theta}, \omega(t)) = \log p(D_{obs}; \hat{\theta}, \omega(t))$$

and

$$\ell_c(\hat{\theta}, \omega(t)) = \log p(D_{com}; \hat{\theta}, \omega(t)).$$

Theoretical Examples

- For the additive $\epsilon-$contamination class of the ITID, we have

$$\mathrm{FI}_B[\mathbf{v}](\boldsymbol{\omega}(0))] = \frac{E\{\sum_{i=1}^{n}[g(\mathbf{d}_{i,c};\boldsymbol{\theta})/p(\mathbf{d}_{i,c};\boldsymbol{\theta}) - 1]|D_{obs}\}^2}{\sum_{i=1}^{n}\mathrm{Var}_S(g(\mathbf{d}_{i,c};\boldsymbol{\theta})/p(\mathbf{d}_{i,c};\boldsymbol{\theta}))},$$

where

$$\mathrm{Var}_S(g(\mathbf{d}_{i,c};\boldsymbol{\theta})/p(\mathbf{d}_{i,c};\boldsymbol{\theta})) = \int [g(\mathbf{d}_{i,c};\boldsymbol{\theta})/p(\mathbf{d}_{i,c};\boldsymbol{\theta}) - 1]^2 p(D_{com},\boldsymbol{\theta};\boldsymbol{\omega}^0)d\Lambda(D_{com},\boldsymbol{\theta}).$$

Theoretical Examples

Cook's posterior mean distance

- $$\mathrm{CM}_h(\omega) = [M_h(\omega) - M_h(\omega^0)]^T C_h [M_h(\omega) - M_h(\omega^0)],$$
  where C_h is chosen to be a positive definite matrix.
- $$M_h(\omega) = \int h(\boldsymbol{\theta}) p(D_{mis}, \boldsymbol{\theta} | D_{obs}; \omega) d\Lambda(D_{mis}, \boldsymbol{\theta}).$$

Theoretical Examples

- We set $f(\boldsymbol{\omega}) = \mathrm{CM}_h(\omega)$ and $\boldsymbol{\omega}(t)$ is a smooth curve on $\mathcal{M}$ with $\boldsymbol{\omega}(0) = \boldsymbol{\omega}^0$ and
$$d_t p_c(\boldsymbol{\omega}(0)) = \mathbf{v} \in T_{\boldsymbol{\omega}(0)}\mathcal{M}.$$
- $\dot{f}(\boldsymbol{\omega}(0)) = 0$ and
$$\ddot{f}(\boldsymbol{\omega}(0)) = \dot{M}_h(\mathbf{v})^T G_h \dot{M}_h(\mathbf{v}),$$
where
$$\dot{M}_h(\mathbf{v}) = \mathrm{Cov}\{h(\boldsymbol{\theta}), d_t \log p(D_{com}, \boldsymbol{\theta}; \boldsymbol{\omega}(t))|D_{obs}\}|_{t=0}. \qquad (27)$$

Theoretical Examples

We consider a simultaneous perturbation to both the prior and the sampling distribution.

- $\mathrm{SI}_{CM_h}[\mathbf{v}](\omega(0)) = \frac{\dot{M}_h(\mathbf{v})^T G_h \dot{M}_h(\mathbf{v})}{g_P(\mathbf{v},\mathbf{v})+g_S(\mathbf{v},\mathbf{v})}$.
- For the perturbation to the prior,

$$p(\boldsymbol{\theta}; t) = p(\boldsymbol{\theta}) + t[g(\boldsymbol{\theta}) - p(\boldsymbol{\theta})],$$

and $\mathrm{SI}_{CM_h}[\mathbf{v}](\omega(0))$ is given by

$$\frac{\mathrm{Cov}\{g(\boldsymbol{\theta})/p(\boldsymbol{\theta}), h(\boldsymbol{\theta})^T|D_{obs}\} C_h \mathrm{Cov}\{h(\boldsymbol{\theta}), g(\boldsymbol{\theta})/p(\boldsymbol{\theta})|D_{obs}\}}{\mathrm{Var}_P(g(\boldsymbol{\theta})/p(\boldsymbol{\theta}))}.$$

Theoretical Examples

- $\mathrm{SI}_{CM_h}[\mathbf{v}](\omega(0))$ is smaller than

$$\mathrm{tr}\left\{\mathrm{Cov}\{h(\boldsymbol{\theta})|D_{obs})\}^{-1}E\left[\left.\frac{\{h(\boldsymbol{\theta})-E[h(\boldsymbol{\theta})|D_{obs}]\}^{\otimes 2}p(\boldsymbol{\theta}|D_{obs})}{p(\boldsymbol{\theta})}\right|D_{obs}\right]\right\}.$$

Simulation Studies

- Data are obtained from N individuals nested within J groups, with group j containing n_j individuals, where $N = \sum_{j=1}^{J} n_j$.
- At level-1, for each group j $(j = 1, \ldots, J)$,

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}_j + \varepsilon_{ij}, \quad i = 1, \ldots, n_j, \tag{28}$$

  where $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.
- At level-2, $\boldsymbol{\beta}_j = \mathbf{Z}_j \boldsymbol{\gamma} + \mathbf{u}_j$, where $\mathbf{u}_j \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.
- The missing data mechanism for y_{ij} is assumed to be missing at random (MAR), and defined as follows:

$$\Pr(r_{ij} = 1 | \mathbf{x}_{ij}, \boldsymbol{\varphi}) = \frac{\exp(\varphi_0 + \boldsymbol{\varphi}_x^T \mathbf{x}_{ij})}{1 + \exp(\varphi_0 + \boldsymbol{\varphi}_x^T \mathbf{x}_{ij})}. \tag{29}$$

Simulation Studies

- We set $J = 100$, $q = 2$, and $r = 3$.
- We varied the values of n_j in order to create a scenario with different cluster sizes. We set $n_1 = \cdots = n_{10} = 3$, $n_{91} = \cdots = n_{100} = 20$, and $n_i \in \{5, 7, 8, 10, 12, 13, 15, 17\}$ for $i = 11, \cdots, 90$.
- We set $\varphi_0 = -2.0$, $\varphi_1 = \varphi_2 = 0.5$, $\boldsymbol{\gamma} = (0.8, 0.8, 0.8)^T$, $\boldsymbol{\Sigma} = 0.5\mathbf{1}_2\mathbf{1}_2^T + 0.5\mathbf{I}_2$ and $\sigma_\varepsilon^2 = 1.0$.
- We independently generated all components (except the intercept) of $\mathbf{x}_{ij}$ and $\mathbf{Z}_j$ from a $U(0, 1)$ distribution.
- $p(\gamma) \stackrel{D}{=} N(\gamma^0, \mathbf{H}_{0\varepsilon})$, $p(\sigma_\varepsilon^{-2}) \stackrel{D}{=} \Gamma(\alpha_{0\varepsilon}, \beta_{0\varepsilon})$, $p(\boldsymbol{\Sigma}) \stackrel{D}{=} IW_q(\rho_0, \mathbf{R}^0)$, where $\gamma^0 = (0.8, 0.8, 0.8)^T$, $\mathbf{R}^0 = 2\mathbf{I}_2 + 2\mathbf{1}_2\mathbf{1}_2^T$, $\alpha_{\varepsilon 0} = 10.0$, $\beta_{\varepsilon 0} = 8.0$, $\rho_0 = 10$, and $\mathbf{H}_{0\varepsilon} = \text{diag}(0.2, 0.2, 0.2)$.

Simulation Studies

Scenario 1: Outlying Clusters

- Generate $\{y_{ij} : j = 1, 99, 100; i = 1, \cdots, n_j\}$ from a normal distribution
$$N(\mathbf{x}_{ij}^T \mathbf{Z}_j \gamma + \mathbf{x}_{ij}^T \mathbf{u}_j, \sigma_\varepsilon^2)$$
with
$$\mathbf{u}_j \sim N_q(5.6\mathbf{1}_2, 1.96\mathbf{I}_2 + 0.3\mathbf{\Sigma}),$$
$(j = 1, 99, 100)$.
- This can be regarded as a case with a wrong distribution for $\mathbf{u}_j$ for $j = 1, 99, 100$.
- We considered a simultaneous perturbation of $\mathbf{u}_j$ and the prior distributions of γ, $\mathbf{\Sigma}$ and σ_ε^2 for the ϕ-divergence.

Simulation Studies

- Simultaneously perturb the distributions of $\mathbf{u}_j$ and the prior distributions of γ, $\mathbf{\Sigma}$ and σ_ε^2.
- No perturbation is $\omega^0 = (1, 1, \ldots, 1, 0)^T$.
- $G(\omega^0) = \text{diag}(G_c(\omega^0), G_\gamma(\omega^0), G_\Sigma(\omega^0), G_\sigma(\omega^0))$ in which $G_c(\omega^0) = q\mathbf{I}_J/2$, $G_\gamma(\omega^0) = r/2$, $G_\Sigma(\omega^0) = \text{Var}_\Sigma[\text{tr}(\mathbf{R}^0\mathbf{\Sigma}^{-1})]/4$ and $G_\sigma(\omega^0) = \text{Var}_{\sigma_\varepsilon^2}[g(\sigma_\varepsilon^{-2})/p(\sigma_\varepsilon^{-2})]$.
- We consider a second scenario with the wrong prior distribution for γ: $p(\gamma) \overset{D}{=} N_2(4\gamma^0, \mathbf{H}_{0\varepsilon})$.

Simulation Studies

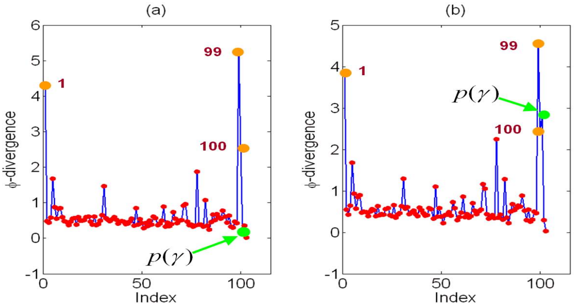

Figure: Group index plots of local influence measures for simultaneous perturbation: (a) $SI_{D_\phi}[\mathbf{e}_j]$ can detect the three influential groups (1, 99, and 100); (b) $SI_{D_\phi}[\mathbf{e}_j]$ can detect both the three influential groups (1, 99, and 100) and the 'incorrect' prior distribution $p(\gamma)$.

Simulation Studies

Scenario 2: Missing-data Mechanism

- Explore the potential deviations of the MAR missing data mechanism in the direction of nonignorable MAR (NMAR).
- We simulated a data set using the same setup as above except that the following missing data mechanism for y_{ij} was assumed,

$$\Pr(r_{ij} = 1 | \mathbf{x}_{ij}, y_{ij}, \boldsymbol{\varphi}, \varphi_y) = \frac{\exp(\varphi_0 + \boldsymbol{\varphi}_x' \mathbf{x}_{ij} + \varphi_y y_{ij})}{1 + \exp(\varphi_0 + \boldsymbol{\varphi}_x' \mathbf{x}_{ij} + \varphi_y y_{ij})} \qquad (30)$$

  with $\varphi_y = 0.5$ to make the missing data fraction approximately 25%.
- When $\varphi_y \neq 0$, the missing mechanism is nonignorable.

Simulation Studies

Sensitivity Analysis: Fix φ_y at a value ω_y to the simulated data set and then vary ω_y in an interval $\Omega_1 = [-2, 2]$.

Table 1. Posterior means (PMs) and standard errors (SDs) of γ at different values of φ_y.

	True $\gamma^0 = (0.8, 0.8, 0.8)$					
	γ_1		γ_2		γ_3	
	PM	SD	PM	SD	PM	SD
$\varphi_y = 0.5$	0.831	0.174	0.721	0.251	0.809	0.255
$\varphi_y = 0.3$	0.777	0.170	0.697	0.249	0.786	0.247
$\varphi_y = 0.15$	0.738	0.167	0.661	0.243	0.776	0.249
$\varphi_y = 0.0$	0.697	0.177	0.622	0.247	0.749	0.250

Simulation Studies

Global Influence Measure

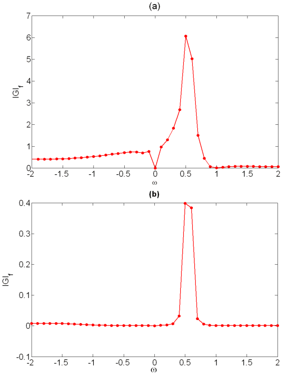

Figure: Plots of $\text{IGI}_f(\omega^0, \omega)$ against $\omega \in \Omega_1$ for (a) $D_\phi(\boldsymbol{\omega})$ and (b) $M_h(\omega)$, in which $h(\theta) = \gamma$.

HIV Data

- A data set from a study of the relationship between acquired immune deficiency syndrome (AIDS) and the use of condoms (Morisky *et al.*, 1998).
- Nine variables about knowledge of AIDS and attitude towards AIDS, belief, and self efficiency of condom use (items 33, 32, 31, 43, 72, 74, 27h, 27e, and 27i in the questionnaire) were taken as manifest variables in $\mathbf{y} = (y_1, \cdots, y_9)'$.
- Variables y_1, y_2, y_3, y_7, y_8 and y_9 were measured via a 5-point scale and hence were treated as continuous; variables y_4, y_5 and y_6 were continuous.
- A continuous item x_1 (item 37) and an ordered categorical item x_2 (item 21, which was treated as continuous) were taken as covariates, x_2 is completely observed.
- 1116 random observations in this data set; the manifest variables and covariates are missing at least once for 361 of them (32.35%).

HIV Data

- $\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\varpi}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \cdots, 1116$, in which $\mathbf{y}_i = (y_{i1}, \cdots, y_{i9})'$ and $\boldsymbol{\varpi}_i = (\eta_i, \xi_{i1}, \xi_{i2})'$ via the following specifications of $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_9)'$ and

$$\boldsymbol{\Lambda}' = \begin{pmatrix} 1.0^* & \lambda_{21} & \lambda_{31} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{52} & \lambda_{62} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{83} & \lambda_{93} \end{pmatrix},$$

  and $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Psi})$ distribution for $i = 1, \cdots, 9$.
- η='threat of AIDS', ξ_1 ='aggressiveness of the sex worker', and ξ_2='worry of contracting AIDS'.
- $\eta_i = b_1 x_{i1} + b_2 x_{i2} + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \delta_i$, where $\delta_i \sim N(0, \psi_\delta)$,
- $\text{logit}\{\Pr(r_{yij} = 1|\boldsymbol{\varphi})\} = \varphi_0 + \varphi_1 y_{i1} + \cdots + \varphi_9 y_{i9}$, where $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \cdots, \varphi_9)'$.
- $\text{logit}\{\Pr(r_{xi1} = 1|\boldsymbol{\varphi}_x)\} = \varphi_{x0} + \omega x_{i1}$.

HIV Data

Global influence measures for the Kullback-Leibler divergence and $M_h(\omega)$

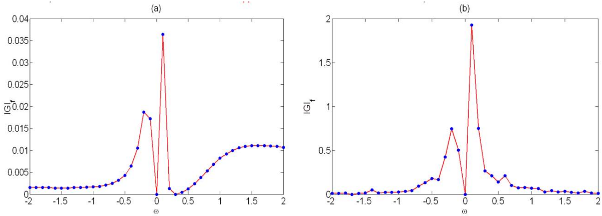

Figure: HIV data: $\mathrm{IGI}_f(\omega^0, \omega)$ against $\omega \in [-2, 2]$ for (a) the Kullback-Leibler divergence and (b) $M_h(\omega)$, in which $h(\theta) = \Gamma = (b_1, b_2, \gamma_1, \gamma_2)^T$.

HIV Data

Sensitivity analysis

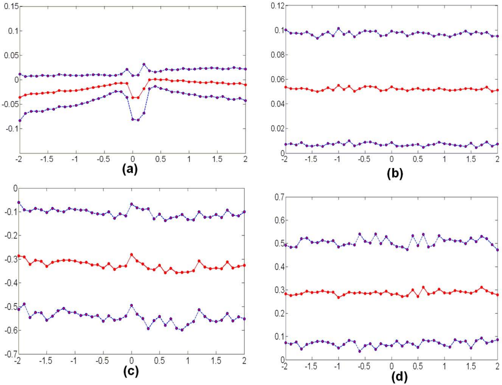

Figure: HIV data: the posterior means (red solid lines and dots) and means$\pm 2\times$SD (blue lines and dots) of b_1 (a), b_2 (b), γ_1 (c), and γ_2 (d) against $\omega \in [-2, 2]$, where SD denotes standard deviation.

HIV Data

- Simultaneous perturbation scheme includes
  - variance perturbation for individual observations
  - perturbation to coefficients in the structural equations model
  - perturbation to
    $\eta_i = b_1 x_{i1} + b_2 x_{i2} + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \omega_{\gamma,1} \xi_{i1}^2 + \omega_{\gamma,2} \xi_{i2}^2 + \omega_{\gamma,3} \xi_{i1} \xi_{i2} + \delta_i$
  - perturbation to the prior distribution of $\boldsymbol{\mu}$
  - perturbation to the prior distribution of $\boldsymbol{\Gamma}$
  - perturbation to the prior distribution of $\boldsymbol{\varphi}$
  - perturbation to $\text{logit}\{\Pr(r_{xi1} = 1 | \boldsymbol{\varphi}_x)\} = \varphi_{x0} + \omega_x x_{i1}$.
- We calculated the local influence measures of the Kullback-Leibler divergence under a simultaneous perturbation scheme.

HIV Data

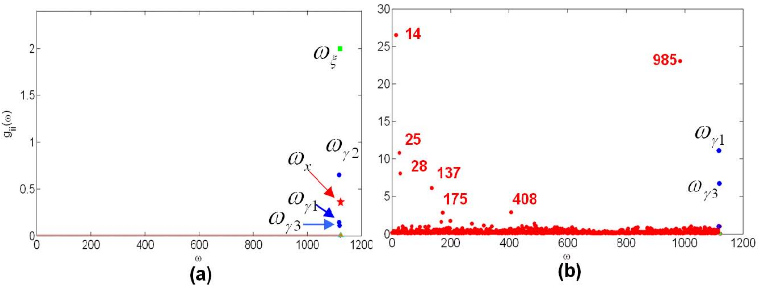

Figure: HIV data: (a) index plot of metric tensor $g_{ii}(\boldsymbol{\omega}^0)/(0.5n)$ for the perturbation (54); (b) Local influence measures $\mathrm{SI}_{\mathrm{D}_\phi}[\mathbf{e}_j]$ for ϕ-divergence.

References

- Zhu, HT., Ibrahim, JG., and Tang, NS. Bayesian influence measures for joint models for longitudinal and survival data. *Biometrics*, in press, 2012.
- Zhu, HT., Ibrahim JG, Tang NS. Bayesian influence approach: a geometric approach. *Biometrika*, 98, 307-323, 2011.
- Ibrahim, J. G., Zhu, H.T., Tang, N. S. Bayesian local influence for survival models (with discussion). *Lifetime Data Analysis*, 17, 43-70, 2011.
- Zhu HT, Ibrahim, JG, Lee SY, and Zhang HP. Appropriate perturbation and influence measures in local influence. *Annals of Statistics*, 35, 2565-2588, 2007.
- Shi, X Y., Zhu HT, Ibrahim, JG. Local influence for generalized linear models with missing covariates. *Biometrics*, 65, 1164-1174, 2009.
- Zhu HT and Zhang HP. A diagnostic procedure based on local influence. *Biometrika*, 91:579-589, 2004.
- Zhu HT and Lee SY. Local influence for generalized linear mixed models. *Canadian Journal of Statistics*, 31:293-309, 2003.
- Zhu HT and Lee SY. Local influence for models with incomplete data. *Journal of the Royal Statistical Society, Series B*, 63:111-126, 2001.

Acknowledgement

- Southeast University: Bocheng Wei
- University of Minnesota: Dennis Cook
- Yale University: Heping Zhang
- University of North Carolina at Chapel Hill: Joseph G. Ibrahim
- Yunnan University: Niansheng Tang
- SAS: Amy Shi