Challenges in Fiducial Inference

Parts of this talk are joint work with
T. C.M Lee (UC Davis), Randy Lai (U of Maine), H. Iyer (NIST)
J. Williams, Y Cui (UNC)

BFF 2017

Jan Hannig[a]

University of North Carolina at Chapel Hill

[a]NSF support acknowledged

Outline

- Introduction
- Definition
- Sparsity
- Regularization
- Conclusions

Outline

- Introduction
- Definition
- Sparsity
- Regularization
- Conclusions

Fiducial?

- Oxford English Dictionary
  - adjective technical (of a point or line) used as a fixed basis of comparison.
  - Origin from Latin fiducia ‘trust, confidence’
- Merriam-Webster dictionary
  1. taken as standard of reference *a fiducial mark*
  2. founded on faith or trust
  3. having the nature of a trust : fiduciary

Aim of this talk

- Explain the definition of generalized fiducial distribution

Aim of this talk

- Explain the definition of generalized fiducial distribution
- Challenge of extra information:
    - Sparsity
    - Regularization

Aim of this talk

- Explain the definition of generalized fiducial distribution
- Challenge of extra information:
  - Sparsity
  - Regularization
- My point of view: frequentist
  - Justified using asymptotic theorems and simulations.
  - GFI tends to work well

Outline

- Introduction
- Definition
- Sparsity
- Regularization
- Conclusions

Comparison to likelihood

- Density is the function $f(\mathbf{x}, \xi)$, where ξ is fixed and $\mathbf{x}$ is variable.

Comparison to likelihood

- Density is the function $f(\mathbf{x}, \xi)$, where ξ is fixed and $\mathbf{x}$ is variable.
- Likelihood is the function $f(\mathbf{x}, \xi)$, where ξ is variable and $\mathbf{x}$ is fixed.

Comparison to likelihood

- Density is the function $f(\mathbf{x}, \xi)$, where ξ is fixed and $\mathbf{x}$ is variable.
- Likelihood is the function $f(\mathbf{x}, \xi)$, where ξ is variable and $\mathbf{x}$ is fixed.
  - Likelihood as a distribution?

General Definition

- Data generating equation $\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{U}, \xi)$.
  - e.g. $X_i = \mu + \sigma U_i$

General Definition

- Data generating equation $\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{U}, \xi)$.
  - e.g. $X_i = \mu + \sigma U_i$
- A distribution on the parameter space is Generalized Fiducial Distribution if it can be obtained as a limit (as $\varepsilon \downarrow 0$) of

$$\arg\min_{\xi} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^{\star}, \xi)\| \mid \{\min_{\xi} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^{\star}, \xi)\| \leq \varepsilon\} \quad (1)$$

General Definition

- Data generating equation $\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{U}, \xi)$.
  - e.g. $X_i = \mu + \sigma U_i$
- A distribution on the parameter space is Generalized Fiducial Distribution if it can be obtained as a limit (as $\varepsilon \downarrow 0$) of

$$\arg\min_{\xi} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^{\star}, \xi)\| \mid \{\min_{\xi} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^{\star}, \xi)\| \leq \varepsilon\} \quad (1)$$

  - Similar to ABC; generating from prior replaced by $\min$.

General Definition

- Data generating equation $\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{U}, \xi)$.
  - e.g. $X_i = \mu + \sigma U_i$
- A distribution on the parameter space is Generalized Fiducial Distribution if it can be obtained as a limit (as $\varepsilon \downarrow 0$) of

$$\arg\min_{\xi} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^{\star}, \xi)\| \mid \{\min_{\xi} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^{\star}, \xi)\| \leq \varepsilon\} \quad (1)$$

  - Similar to ABC; generating from prior replaced by $\min$.
  - Is this practicle? Can we compute?

Explicit limit (1)

- Assume $\mathbf{X} \in \mathbb{R}^n$ is continuous; parameter $\xi \in \mathbb{R}^p$
- The limit in (1) has density (H, Iyer, Lai & Lee, 2016)

$$r(\xi|\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x}|\xi) J(\mathbf{x},\xi)}{\int_{\Xi} f_{\mathbf{X}}(\mathbf{x}|\xi') J(\mathbf{x},\xi')\, d\xi'},$$

where $J(\mathbf{x},\xi) = D\left(\frac{d}{d\xi} \mathbf{G}(\mathbf{u},\xi) \Big|_{\mathbf{u}=\mathbf{G}^{-1}(\mathbf{x},\xi)} \right)$

- $n = p$ gives $D(A) = |\det A|$

Explicit limit (1)

- Assume $\mathbf{X} \in \mathbb{R}^n$ is continuous; parameter $\xi \in \mathbb{R}^p$
- The limit in (1) has density (H, Iyer, Lai & Lee, 2016)

$$r(\xi|\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x}|\xi)J(\mathbf{x},\xi)}{\int_{\Xi} f_{\mathbf{X}}(\mathbf{x}|\xi')J(\mathbf{x},\xi')\,d\xi'},$$

where $J(\mathbf{x},\xi) = D\left(\frac{d}{d\xi}\mathbf{G}(\mathbf{u},\xi)\Big|_{\mathbf{u}=\mathbf{G}^{-1}(\mathbf{x},\xi)}\right)$

- $n = p$ gives $D(A) = |\det A|$
- $\|\cdot\|_2$ gives $D(A) = (\det A^\top A)^{1/2}$
  Compare to Fraser, Reid, Marras & Yi (2010)
- $\|\cdot\|_\infty$ gives $D(A) = \sum_{\mathbf{i}=(i_1,\ldots,i_p)} |\det(A)_{\mathbf{i}}|$

Example -- Linear Regression

- Data generating equation $Y = X\beta + \sigma Z$

Example -- Linear Regression

- Data generating equation $Y = X\beta + \sigma Z$
- $\frac{d}{d\theta} Y = (X, Z)$ and $Z = (Y - X\beta)/\sigma$.
- The L_2 Jacobian is

$$J(\boldsymbol{y}, \beta, \sigma) = \left(\det\left((\boldsymbol{X}, \frac{\boldsymbol{y} - \boldsymbol{X}\beta}{\sigma})^\top (\boldsymbol{X}, \frac{\boldsymbol{y} - \boldsymbol{X}\beta}{\sigma})\right)\right)^{1/2}$$

$$= \sigma^{-1} |\det(X^T X)|^{1/2} (RSS)^{1/2}$$

Example -- Linear Regression

- Data generating equation $Y = X\beta + \sigma Z$
- $\frac{d}{d\theta} Y = (X, Z)$ and $Z = (Y - X\beta)/\sigma$.
- The L_2 Jacobian is

$$J(\boldsymbol{y}, \beta, \sigma) = \left(\det\left((\boldsymbol{X}, \frac{\boldsymbol{y} - \boldsymbol{X}\beta}{\sigma})^\top (\boldsymbol{X}, \frac{\boldsymbol{y} - \boldsymbol{X}\beta}{\sigma})\right)\right)^{1/2}$$
$$= \sigma^{-1} |\det(X^T X)|^{1/2} (RSS)^{1/2}$$

- Fiducial *happens to be* same as independence Jeffreys, *explicit* normalizing constant

Example -- Uniform(θ, θ^2)

- X_i i.i.d. $U(\theta, \theta^2)$, $\theta > 1$

Example -- Uniform(θ, θ^2)

- X_i i.i.d. $U(\theta, \theta^2)$, $\theta > 1$
  - Data generating equation $X_i = \theta + (\theta^2 - \theta)U_i,\ U_i \sim U(0,1)$.

Example -- Uniform(θ, θ^2)

- X_i i.i.d. $U(\theta, \theta^2)$, $\theta > 1$
  - Data generating equation $X_i = \theta + (\theta^2 - \theta)U_i,\ U_i \sim U(0,1)$.
- Compute Jacobian: $\frac{d}{d\theta}[\theta + (\theta^2 - \theta)U_i] = 1 + (2\theta - 1)U_i$, with $U_i = \frac{X_i - \theta}{\theta^2 - \theta}$.

Example -- Uniform(θ, θ^2)

- X_i i.i.d. $U(\theta, \theta^2)$, $\theta > 1$
  - Data generating equation $X_i = \theta + (\theta^2 - \theta)U_i,\ U_i \sim U(0,1)$.
- Compute Jacobian: $\frac{d}{d\theta}[\theta + (\theta^2 - \theta)U_i] = 1 + (2\theta - 1)U_i$, with $U_i = \frac{X_i - \theta}{\theta^2 - \theta}$.
  - Using $\|.\|_\infty$ we have $J(\mathbf{x}, \theta) = n\frac{\bar{x}(2\theta-1)-\theta^2}{\theta^2-\theta}$.

Example -- Uniform(θ, θ^2)

- X_i i.i.d. $U(\theta, \theta^2)$, $\theta > 1$
  - Data generating equation $X_i = \theta + (\theta^2 - \theta)U_i,\ U_i \sim U(0,1)$.
- Compute Jacobian: $\frac{d}{d\theta}[\theta + (\theta^2 - \theta)U_i] = 1 + (2\theta - 1)U_i$, with $U_i = \frac{X_i - \theta}{\theta^2 - \theta}$.
  - Using $\|\cdot\|_\infty$ we have $J(\mathbf{x}, \theta) = n\frac{\bar{x}(2\theta-1)-\theta^2}{\theta^2-\theta}$.
- Reference prior $\pi(\theta) = \frac{e^{\psi\left(\frac{2\theta}{2\theta-1}\right)}(2\theta-1)}{\theta(\theta-1)}$ Berger, Bernardo & Sun (2009) – complicated to derive.

Example -- Uniform(θ, θ^2)

- X_i i.i.d. $U(\theta, \theta^2)$, $\theta > 1$
  - Data generating equation $X_i = \theta + (\theta^2 - \theta)U_i,\ U_i \sim U(0,1)$.
- Compute Jacobian: $\frac{d}{d\theta}[\theta + (\theta^2 - \theta)U_i] = 1 + (2\theta - 1)U_i$, with $U_i = \frac{X_i - \theta}{\theta^2 - \theta}$.
  - Using $\|\cdot\|_\infty$ we have $J(\mathbf{x}, \theta) = n\frac{\bar{x}(2\theta-1)-\theta^2}{\theta^2-\theta}$.
- Reference prior $\pi(\theta) = \frac{e^{\psi\left(\frac{2\theta}{2\theta-1}\right)}(2\theta-1)}{\theta(\theta-1)}$ Berger, Bernardo & Sun (2009) – complicated to derive.
  - In simulations fiducial was marginally better than reference prior which was much better than flat prior.

Important Simple Observations

- GFD is allways proper

Important Simple Observations

- GFD is allways proper
- GFD is invariant to re-parametrizations (same as Jeffreys)

Important Simple Observations

- GFD is allways proper
- GFD is invariant to re-parametrizations (same as Jeffreys)
- GFD is *not* invariant to smooth transformation of the data if $n > p$

Important Simple Observations

- GFD is allways proper
- GFD is invariant to re-parametrizations (same as Jeffreys)
- GFD is *not* invariant to smooth transformation of the data if $n > p$
- Does not satisfy likelihood principle.

Various Asymptotic Results

$r(\xi|\mathbf{x}) \propto f_{\mathbf{X}}(\mathbf{x}|\xi) J(\mathbf{x},\xi)$ where $J(\mathbf{x},\xi) = D\left(\frac{d}{d\xi}\mathbf{G}(\mathbf{u},\xi)\Big|_{\mathbf{u}=\mathbf{G}^{-1}(\mathbf{x},\xi)} \right)$

- Most start with $C_n^{-1} J(\mathbf{x},\xi) \to J(\xi_0,\xi)$
- Bernstein-von Mises theorem for fiducial distributions provides asymptotic correctness of fiducial CIs H (2009, 2013), Sonderegger & H (2013) .
- Consistency of model selection H & Lee (2009), Lai, H & Lee (2015), H, Iyer, Lai & Lee (2016).
- Regular higher order asymptotics in Pal Majumdar & H (2016+).

Outline

- Introduction
- Definition
- Sparsity
- Regularization
- Conclusions

Model Selection

- $\boldsymbol{X} = \boldsymbol{G}(M, \boldsymbol{\xi}_M, \boldsymbol{U}), \qquad M \in \mathcal{M},\ \boldsymbol{\xi}_M \in \boldsymbol{\xi}_M$

Theorem: (H, Iyer, Lai, Lee 2016) Under assumptions

$$r(M|\boldsymbol{y}) \propto q^{|M|} \int_{\boldsymbol{\xi}_M} f_M(\boldsymbol{y}, \boldsymbol{\xi}_M) J_M(\boldsymbol{y}, \boldsymbol{\xi}_M)\, d\boldsymbol{\xi}_M$$

Model Selection

- $\boldsymbol{X} = \boldsymbol{G}(M, \boldsymbol{\xi}_M, \boldsymbol{U}), \qquad M \in \mathcal{M},\ \boldsymbol{\xi}_M \in \boldsymbol{\Xi}_M$

Theorem: (H, Iyer, Lai, Lee 2016) Under assumptions

$$r(M|\boldsymbol{y}) \propto q^{|M|} \int_{\boldsymbol{\Xi}_M} f_M(\boldsymbol{y}, \boldsymbol{\xi}_M) J_M(\boldsymbol{y}, \boldsymbol{\xi}_M)\, d\boldsymbol{\xi}_M$$

- Need for penalty – in fiducial framework additional equations $0 = P_k, \quad k = 1, \ldots, \min(|M|, n)$

Model Selection

- $\boldsymbol{X} = \boldsymbol{G}(M, \boldsymbol{\xi}_M, \boldsymbol{U}), \qquad M \in \mathcal{M},\ \boldsymbol{\xi}_M \in \boldsymbol{\xi}_M$

Theorem: (H, Iyer, Lai, Lee 2016) Under assumptions

$$r(M|\boldsymbol{y}) \propto q^{|M|} \int_{\boldsymbol{\xi}_M} f_M(\boldsymbol{y}, \boldsymbol{\xi}_M) J_M(\boldsymbol{y}, \boldsymbol{\xi}_M)\, d\boldsymbol{\xi}_M$$

- Need for penalty – in fiducial framework additional equations $0 = P_k, \quad k = 1, \ldots, \min(|M|, n)$
  - Default value $q = n^{-1/2}$ (motivated by MDL)

Alternative to penalty

- Penalty is used to discourage models with many parameters

Alternative to penalty

- Penalty is used to discourage models with many parameters
- Real issue: Not too many parameters but a smaller model can do almost the same job

Alternative to penalty

- Penalty is used to discourage models with many parameters
- Real issue: Not too many parameters but a smaller model can do almost the same job

$$r(M|\boldsymbol{y}) \propto \int_{\boldsymbol{\xi}_M} f_M(\boldsymbol{y}, \boldsymbol{\xi}_M) J_M(\boldsymbol{y}, \boldsymbol{\xi}_M) h_M(\boldsymbol{\xi}_M)\, d\boldsymbol{\xi}_M,$$

$$h_M(\boldsymbol{\xi}_M) = \begin{cases} 0 & \text{a smaller model predicts nearly as well} \\ 1 & \text{otherwise} \end{cases}$$

Alternative to penalty

- Penalty is used to discourage models with many parameters
- Real issue: Not too many parameters but a smaller model can do almost the same job

$$r(M|\boldsymbol{y}) \propto \int_{\boldsymbol{\xi}_M} f_M(\boldsymbol{y}, \boldsymbol{\xi}_M) J_M(\boldsymbol{y}, \boldsymbol{\xi}_M) h_M(\boldsymbol{\xi}_M)\, d\boldsymbol{\xi}_M,$$

$$h_M(\boldsymbol{\xi}_M) = \begin{cases} 0 & \text{a smaller model predicts nearly as well} \\ 1 & \text{otherwise} \end{cases}$$

- Motivated by non-local priors of Johnson & Rossell (2009)

Regression

- $\boldsymbol{Y} = X\beta + \sigma Z$
- First idea $h_M(\beta_M) = I_{\{|\beta_i|>\epsilon,\ i\in M\}}$ – issue: collinearity

Regression

- $\boldsymbol{Y} = X\beta + \sigma Z$
- First idea $h_M(\beta_M) = I_{\{|\beta_i| > \epsilon,\ i \in M\}}$ – issue: collinearity
- Better:

$$h_M(\beta_M) := I_{\left\{\frac{1}{2}\|X^T(X_M\beta_M - Xb_{min})\|_2^2 \geq \varepsilon(n, |M|)\right\}}$$

where b_{min} solves

$$\min_{b \in R^p} \frac{1}{2}\|X^T(X_M\beta_M - Xb)\|_2^2 \quad \text{subject to} \quad \|b\|_0 \leq |M| - 1.$$

  - algorithm – Bertsimas et al (2016)

Regression

- $\boldsymbol{Y} = X\beta + \sigma Z$
- First idea $h_M(\beta_M) = I_{\{|\beta_i|>\epsilon,\ i \in M\}}$ – issue: collinearity
- Better:

$$h_M(\beta_M) := I_{\left\{\frac{1}{2}\|X^T(X_M\beta_M - Xb_{min})\|_2^2 \geq \varepsilon(n,|M|)\right\}}$$

where b_{min} solves

$$\min_{b \in R^p} \frac{1}{2}\|X^T(X_M\beta_M - Xb)\|_2^2 \quad \text{subject to} \quad \|b\|_0 \leq |M| - 1.$$

  - algorithm – Bertsimas et al (2016)
  - similar to Dantzig selector Candes & Tao (2007) different norm and target

GFD

$$r(M|\boldsymbol{y}) \propto \pi^{\frac{|M|}{2}} \Gamma\left(\frac{n-|M|}{2}\right) RSS_M^{-\left(\frac{n-|M|-1}{2}\right)} E[h_M^{\varepsilon}(\beta_M^{\star})]$$

Observations:

- Expectation with respect to within model GFD (usual T)

GFD

$$r(M|\boldsymbol{y}) \propto \pi^{\frac{|M|}{2}} \Gamma\left(\frac{n-|M|}{2}\right) RSS_M^{-\left(\frac{n-|M|-1}{2}\right)} E[h_M^{\varepsilon}(\beta_M^{\star})]$$

Observations:

- Expectation with respect to within model GFD (usual T)
- $r(M|y)$ negligibly small for large models because of h, e.g., $|M| > n$ implies $r(M|y) = 0$.

GFD

$$r(M|\boldsymbol{y}) \propto \pi^{\frac{|M|}{2}} \Gamma\left(\frac{n-|M|}{2}\right) RSS_M^{-\left(\frac{n-|M|-1}{2}\right)} E[h_M^{\varepsilon}(\beta_M^{\star})]$$

Observations:

- Expectation with respect to within model GFD (usual T)
- $r(M|y)$ negligibly small for large models because of h, e.g., $|M| > n$ implies $r(M|y) = 0$.
- Implemented using Grouped Independence Metropolis Hastings (Andrieu & Roberts, 2009).

Main Result

Theorem Williams & H (2017+)
Suppose the true model is given by M_T. Then under certain conditions, for a fixed positive constant $\alpha < 1$,

$$r(M_T|y) = \frac{r(M_T|y)}{\sum_{j=1}^{n^\alpha} \sum_{M:|M|=j} r(M|y)} \xrightarrow{P} 1 \text{ as } n, p \to \infty.$$

Some Conditions

- Number of Predictors: $\liminf_{\substack{n\to\infty \\ p\to\infty}} \frac{n^{1-\alpha}}{\log(p)} > 2$,

Some Conditions

- Number of Predictors: $\liminf_{\substack{n\to\infty \\ p\to\infty}} \frac{n^{1-\alpha}}{\log(p)} > 2$,
- For the true model/parameter $p_T < \log n^{\gamma}$

$$\varepsilon_{M_T}(n,p) \leq \frac{1}{18} \|X^T(\mu_T - X b_{min})\|_2^2$$

where b_{min} minimizes the norm subject to $\|b\|_0 \leq p_T - 1$.

Some Conditions

- Number of Predictors: $\liminf_{\substack{n\to\infty \\ p\to\infty}} \frac{n^{1-\alpha}}{\log(p)} > 2$,
- For the true model/parameter $p_T < \log n^\gamma$

$$\varepsilon_{M_T}(n,p) \leq \frac{1}{18}\|X^T(\mu_T - Xb_{min})\|_2^2$$

where b_{min} minimizes the norm subject to $\|b\|_0 \leq p_T - 1$.

- For a large model $|M| > p_T$ and large enough n or p,

$$\frac{9}{2}\|X^T(H_M - H_{M(-1)})\mu_T\|_2^2 < \varepsilon_M(n,p),$$

where H_M and $H_{M(-1)}$ are the projection matrix for M and M with a covariate removed respectively.

Simulation

- Setup from Rockova & George (2015)
  - $n = 100$, $p = 1000$, $p_T = 8$.
  - Columns of X either a) independent or b) correlated with $\rho = 0.6$
  - $\varepsilon_M(n,p) = \Lambda_M \widehat{\sigma}^2_M \left(\frac{n^{0.51}}{9} + |M| \frac{\log(p\pi)^{1.1}}{9} - \log(n)^{\gamma} \right)_+$ with $\gamma = 1.45$.

Highlight of simulation results

- See Jon Williams' poster for details on theory and simulation

Highlight of simulation results

- See Jon Williams' poster for details on theory and simulation
- When X independent – usually select the correct model
- When X correlated – usually select too small of a model

Highlight of simulation results

- See Jon Williams' poster for details on theory and simulation
- When X independent – usually select the correct model
- When X correlated – usually select too small of a model
  - Conditions of Theorem violated

Highlight of simulation results

- See Jon Williams' poster for details on theory and simulation
- When X independent – usually select the correct model
- When X correlated – usually select too small of a model
  - Conditions of Theorem violated
  - based on conditions: p decreased to 500 to satisfy, performance improves.

Comments

Comments

- Standardized way of measuring closeness in other models?

Comments

- Standardized way of measuring closeness in other models?
- What if small model not the right target, e.g., gene interactions?

Outline

- Introduction
- Definition
- Sparsity
- Regularization
- Conclusions

Recall

- A distribution on the parameter space is Generalized Fiducial Distribution if it can be obtained as a limit (as $\varepsilon \downarrow 0$) of

$$\arg\min_{\xi} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^{\star}, \xi)\| \mid \{\min_{\xi} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^{\star}, \xi)\| \leq \varepsilon\}$$

Recall

- A distribution on the parameter space is Generalized Fiducial Distribution if it can be obtained as a limit (as $\varepsilon \downarrow 0$) of

$$\arg\min_{\xi} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^{\star}, \xi)\| \mid \{\min_{\xi} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^{\star}, \xi)\| \leq \varepsilon\}$$

- Conditioning $U^{\star}$ on $\{\mathbf{x} = \boldsymbol{G}(\mathbf{U}^{\star}, \xi)\}$

  – “regularization by model”

Most general iid model

- Data generating equation:

$$X_i = F^{-1}(U_i), \quad U_i, \text{ i.i.d. Uniform}(0,1)$$

Most general iid model

- Data generating equation:

$$X_i = F^{-1}(U_i), \quad U_i, \text{ i.i.d. Uniform}(0,1)$$

- Inverting (solving for F) we get

$$F^*(x_i^-) \leq U_i^* \leq F^*(x_i).$$

There is a solution iff order of U_i^* matches order of x_i.

Most general iid model

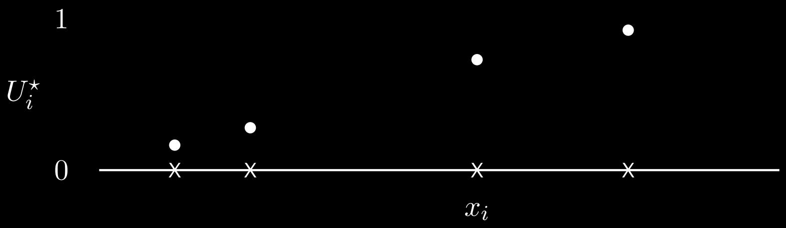
- Data generating equation:

$$X_i = F^{-1}(U_i), \quad U_i, \text{ i.i.d. Uniform}(0,1)$$

- Inverting (solving for F) we get

$$F^*(x_i^-) \leq U_i^* \leq F^*(x_i).$$

There is a solution iff order of U_i^* matches order of x_i.



Most general iid model

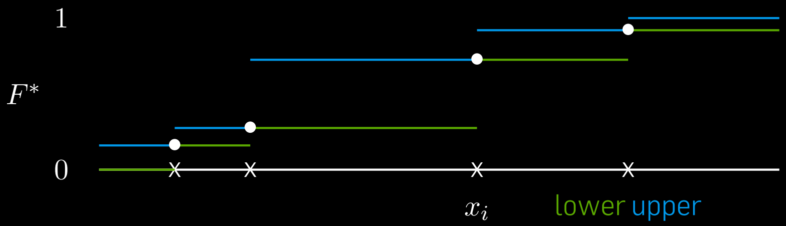
- Data generating equation:

$$X_i = F^{-1}(U_i), \quad U_i, \text{ i.i.d. Uniform}(0,1)$$

- Inverting (solving for F) we get

$$F^*(x_i^-) \leq U_i^* \leq F^*(x_i).$$

There is a solution iff order of U_i^* matches order of x_i.



Most general iid model

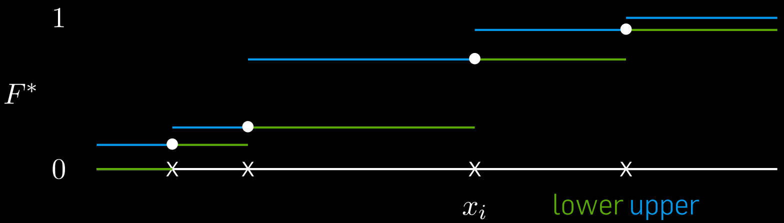
- Data generating equation:

$$X_i = F^{-1}(U_i), \quad U_i, \text{ i.i.d. Uniform}(0,1)$$

- Inverting (solving for F) we get

$$F^*(x_i^-) \le U_i^* \le F^*(x_i).$$

There is a solution iff order of U_i^* matches order of x_i.



- See Yifan Cui's poster for extension to censored data.

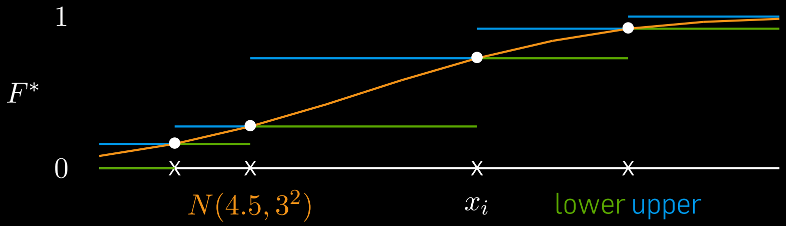
Additional Constraints

- Location scale family with known density $f(x)$ and cdf $F(x)$, e.g., $N(\mu, \sigma^2)$.

Additional Constraints

- Location scale family with known density $f(x)$ and cdf $F(x)$, e.g., $N(\mu, \sigma^2)$.
- Condition U_i^* on existence μ^*, σ^* so that

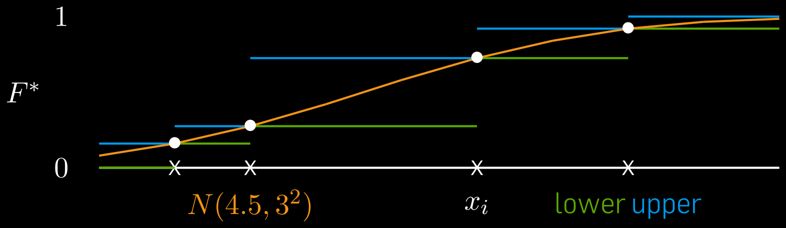
$$F(\sigma^{*-1}(x_i - \mu^*)) = U_i^*, \quad \text{for all } i$$



Additional Constraints

- Location scale family with known density $f(x)$ and cdf $F(x)$, e.g., $N(\mu, \sigma^2)$.
- Condition U_i^* on existence μ^*, σ^* so that

$$F(\sigma^{*-1}(x_i - \mu^*)) = U_i^*, \quad \text{for all } i$$



- GFD is $r(\mu, \sigma) \propto \sigma^{-1} \prod_{i=1}^n \sigma^{-1} f\left(\sigma^{-1}(x_i - \mu)\right)$

Constraint complications

Toy example: $X = \mu + Z, \quad \mu > 0$.

Constraint complications

Toy example: $X = \mu + Z, \quad \mu > 0$.

- Option 1: condition $Z^\star | x - Z^\star > 0$
  - $r(\mu) = \frac{\varphi(x-\mu)}{\Phi(x)} I_{\{\mu>0\}}$
  - Lower confidence bounds do not have correct coverage.

Constraint complications

Toy example: $X = \mu + Z, \quad \mu > 0$.

- Option 1: condition $Z^\star | x - Z^\star > 0$
  - $r(\mu) = \frac{\varphi(x-\mu)}{\Phi(x)} I_{\{\mu>0\}}$
  - Lower confidence bounds do not have correct coverage.
- Option 2: projection to $\mu > 0$
  - $r(\mu) = (1 - \Phi(x)) I_{\{0\}} + \varphi(x - \mu) I_{\{\mu>0\}}$
  - Correct coverage; possible to get $\{0\}$ as CI – sure bet against

Constraint complications

Toy example: $X = \mu + Z, \quad \mu > 0$.

- Option 1: condition $Z^\star | x - Z^\star > 0$
  - $r(\mu) = \frac{\varphi(x-\mu)}{\Phi(x)} I_{\{\mu>0\}}$
  - Lower confidence bounds do not have correct coverage.
- Option 2: projection to $\mu > 0$
  - $r(\mu) = (1 - \Phi(x)) I_{\{0\}} + \varphi(x - \mu) I_{\{\mu>0\}}$
  - Correct coverage; possible to get $\{0\}$ as CI – sure bet against
- Option 3: mixture
  - $r(\mu) = \min(\frac{1}{2}, 1 - \Phi(x))) I_{\{0\}} + \max(\frac{1}{2\Phi(x)}, 1) \varphi(x - \mu) I_{\{\mu>0\}}$
  - Correct/conservative coverage, no $\{0\}$ for reasonable α CIs.

Shape restrictions - preliminary results

- Example: Positive iid data with concave cdf
  (MLE is the Grenander estimator)

Shape restrictions - preliminary results

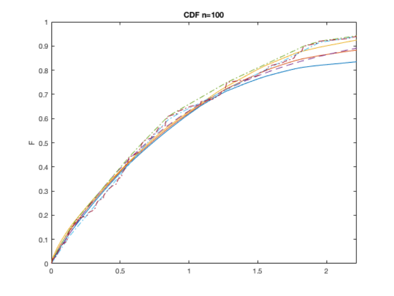
- Example: Positive iid data with concave cdf
  (MLE is the Grenander estimator)
  - Condition $U^\star$ on concave solution (Gibbs sampler)

Shape restrictions - preliminary results

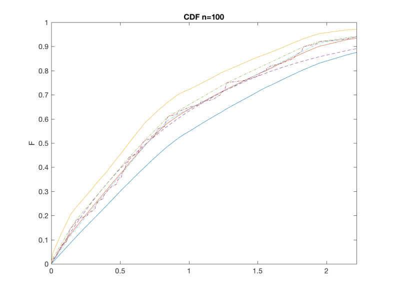
- Example: Positive iid data with concave cdf
  (MLE is the Grenander estimator)
  - Condition $U^\star$ on concave solution (Gibbs sampler)
  - Project unrestricted GFD to space of concave functions (quadratic program)

Shape restrictions - preliminary results

- Example: Positive iid data with concave cdf
  (MLE is the Grenander estimator)
  - Condition $U^\star$ on concave solution (Gibbs sampler)
  - Project unrestricted GFD to space of concave functions (quadratic program)



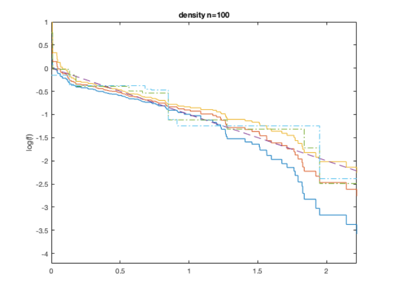
Condition



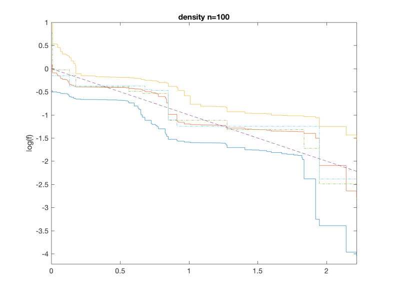
Projection

Shape restrictions - preliminary results

- Example: Positive iid data with concave cdf (MLE is the Grenander estimator)
  - Condition $U^\star$ on concave solution (Gibbs sampler)
  - Project unrestricted GFD to space of concave functions (quadratic program)



Condition



Projection

Comments

- When to use conditioning vs. projection?

Comments

- When to use conditioning vs. projection?
- Connection to ancillarity and IM.

Comments

- When to use conditioning vs. projection?
- Connection to ancillarity and IM.
- Computational cost a consideration?

Outline

Fiducial Future

Fiducial Future

- What is it that we provide?
  - GFI: General purpose method that often works well

Fiducial Future

- What is it that we provide?
  - GFI: General purpose method that often works well
- Computational convenience and efficiency
  - Fiducial options in software.

Fiducial Future

- What is it that we provide?
  - GFI: General purpose method that often works well
- Computational convenience and efficiency
  - Fiducial options in software.
- Theoretical guarantees

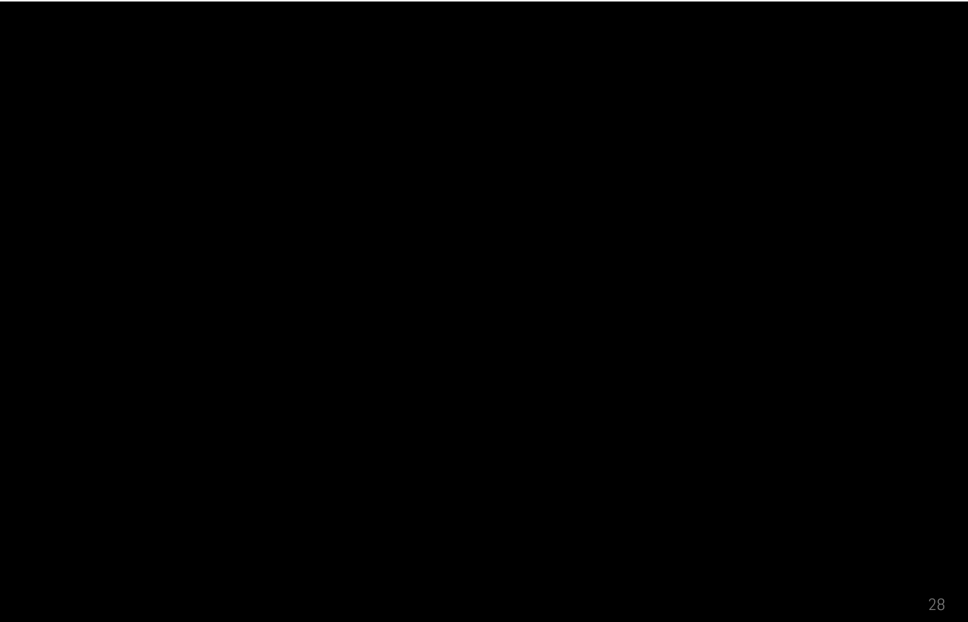
Fiducial Future

- What is it that we provide?
  - GFI: General purpose method that often works well
- Computational convenience and efficiency
  - Fiducial options in software.
- Theoretical guarantees
- Applications
  - The proof is in the pudding

List of successful applications

- General Linear Mixed Models E, H & Iyer (2008); Cissewski & H (2012)
- Confidence sets for wavelet regression H & Lee (2009) and free knot splines Sonderegger & H (2014)
- Extreme value data (Generalized Pareto), Maximum mean, and model comparison Wandler & H (2011, 2012ab)
- Uncertainty quantification for ultra high dimensional regression Lai, H & Lee (2015), Wandler & H (2017+)
- Volatility estimation for high frequency data Katsoridas & H (2016+)
- Logistic regression with random effects (response models) Liu & H (2016,2017)

I have a dream …



I have a dream ...

- One famous statistician said (I paraphrase)

  "I use Bayes because there is no need to prove asymptotic theorem; it is correct."

I have a dream ...

- One famous statistician said (I paraphrase)

  "I use Bayes because there is no need to prove asymptotic theorem; it is correct."

- I have a dream that by the time I retire people will have similar trust in fiducial inspired approaches.

I have a dream ...

- One famous statistician said (I paraphrase)

  "I use Bayes because there is no need to prove asymptotic theorem; it is correct."

- I have a dream that by the time I retire people will have similar trust in fiducial inspired approaches.

Thank you!