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\textsuperscript{a}

University of North Carolina at Chapel Hill

\textsuperscript{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 $f(x, \xi)$, where $\xi$ is fixed and $x$ is variable.
Comparison to likelihood

- **Density** is the function $f(x, \xi)$, where $\xi$ is fixed and $x$ is variable.
- **Likelihood** is the function $f(x, \xi)$, where $\xi$ is variable and $x$ is fixed.
Comparison to likelihood

- **Density** is the function $f(x, \xi)$, where $\xi$ is fixed and $x$ is variable.
- **Likelihood** is the function $f(x, \xi)$, where $\xi$ is variable and $x$ is fixed.
  - Likelihood as a distribution?
General Definition

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

- Data generating equation $X = G(U, \xi)$.
  - 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

$$\arg\min_{\xi} \|x - G(U^*, \xi)\| \mid \{\min_{\xi} \|x - G(U^*, \xi)\| \leq \varepsilon\}$$  \hspace{1cm} (1)
General Definition

- Data generating equation $\mathbf{X} = G(U, \xi)$.
  - 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

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

- Similar to ABC; generating from prior replaced by $\min$. 
Data generating equation $X = G(U, \xi)$.
- 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

$$\arg \min_{\xi} \| x - G(U^*, \xi) \| \mid \{ \min_{\xi} \| x - G(U^*, \xi) \| \leq \varepsilon \} \tag{1}$$

- Similar to ABC; generating from prior replaced by min.
- Is this practical? 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 \((H, Iyer, Lai & Lee, 2016)\)

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

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

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

- $n = p$ 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 $D(A) = \sum_{i=(i_1, \ldots, i_p)} |\det(A)_i|$
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

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

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

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

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

- Fiducial happens to be same as independence Jeffreys, explicit normalizing constant
Example -- Uniform($\theta, \theta^2$)

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

- $X_i$ 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($\theta, \theta^2$)

- $X_i$ 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: $\frac{\frac{d}{d\theta} [\theta + (\theta^2 - \theta)U_i]}{\frac{X_i - \theta}{\theta^2 - \theta}} = 1 + (2\theta - 1)U_i$, with $U_i = \frac{X_i - \theta}{\theta^2 - \theta}$. 

Using $\|\|$ we have $J(x; \theta) = n(2\theta - 1)^2$.

Reference prior $(\cdot) = e^{(2\theta - 1)}(2\theta - 1)(\cdot)$.


In simulations fiducial was marginally better than reference prior which was much better than flat prior.
Example -- Uniform($\theta, \theta^2$)

- $X_i$ 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 $J(x, \theta) = n \bar{x} \frac{(2\theta - 1) - \theta^2}{\theta^2 - \theta}$.
Example -- Uniform($\theta, \theta^2$)

- $X_i$ 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: $\frac{d}{d\theta} \left[ \theta + (\theta^2 - \theta)U_i \right] = 1 + (2\theta - 1)U_i$, with $U_i = \frac{X_i - \theta}{\theta^2 - \theta}$.

  - Using $\| \cdot \|_\infty$ we have $J(x, \theta) = n\frac{\bar{x}(2\theta - 1) - \theta^2}{\theta^2 - \theta}$.

- Reference prior $\pi(\theta) = \frac{e^{\psi(\frac{2\theta}{2\theta - 1})(2\theta - 1)}}{\theta(\theta - 1)}$ Berger, Bernardo & Sun (2009) – complicated to derive.
Example -- Uniform(θ, θ²)

- $X_i$ i.i.d. $U(θ, θ^2)$, $θ > 1$
  - Data generating equation: $X_i = θ + (θ^2 - θ)U_i$, $U_i \sim U(0, 1)$.
- Compute Jacobian: $\frac{d}{dθ}[θ + (θ^2 - θ)U_i] = 1 + (2θ - 1)U_i$, with $U_i = \frac{X_i - θ}{θ^2 - θ}$.
  - Using $\| \cdot \|_\infty$ we have $J(x, θ) = n\bar{x}(2θ - 1) - θ^2$.
- Reference prior $π(θ) = \frac{e^{\psi\left(\frac{2θ}{2θ-1}\right)}(2θ-1)}{θ(θ-1)}$ 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 always proper
Important Simple Observations

- GFD is always proper
- GFD is invariant to re-parametrizations (same as Jeffreys)
Important Simple Observations

- GFD is always proper
- GFD is invariant to re-parametrizations (same as Jeffreys)
- GFD is \textit{not} invariant to smooth transformation of the data if $n > p$
Important Simple Observations

- GFD is always 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

\[ r(\xi|x) \propto f_X(x|\xi) J(x, \xi) \text{ where } J(x, \xi) = D \left( \frac{d}{d\xi} G(u, \xi) \bigg|_{u=G^{-1}(x, \xi)} \right) \]

- Most start with \( C_n^{-1} J(x, \xi) \rightarrow J(\xi_0, \xi) \)
- Regular higher order asymptotics in Pal Majumdar & H (2016+).
Outline

- Introduction
- Definition
- Sparsity
- Regularization
- Conclusions
Model Selection

\[ X = G(M, \xi_M, U), \quad M \in \mathcal{M}, \xi_M \in \xi_M \]

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

\[ r(M|y) \propto q^M \int_{\xi_M} f_M(y, \xi_M) J_M(y, \xi_M) \, d\xi_M \]
Model Selection

- \( \mathbf{X} = \mathbf{G}(M, \xi_M, \mathbf{U}), \quad M \in \mathcal{M}, \xi_M \in \xi_M \)

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

\[
\begin{align*}
r(M|\mathbf{y}) \propto q^{|M|} \int_{\xi_M} f_M(\mathbf{y}, \xi_M) J_M(\mathbf{y}, \xi_M) d\xi_M
\end{align*}
\]

- Need for penalty – in fiducial framework additional equations

\[
0 = P_k, \quad k = 1, \ldots, \min(|M|, n)
\]
Model Selection

\[ X = G(M, \xi_M, U), \quad M \in \mathcal{M}, \ \xi_M \in \xi_M \]

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

\[ r(M|y) \propto q^{|M|} \int_{\xi_M} f_M(y, \xi_M) J_M(y, \xi_M) \, d\xi_M \]

- 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

\[
r(M|y) \propto \int_{\xi_M} f_M(y, \xi_M) J_M(y, \xi_M) h_M(\xi_M) d\xi_M,
\]

\[
h_M(\xi_M) = \begin{cases} 
0 & \text{a smaller model predicts nearly as well} \\
1 & \text{otherwise}
\end{cases}
\]
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

\[
\begin{align*}
    r(M|y) &\propto \int_{\xi_M} f_M(y, \xi_M) J_M(y, \xi_M) h_M(\xi_M) d\xi_M, \\
    h_M(\xi_M) &= \begin{cases} 
    0 & \text{a smaller model predicts nearly as well} \\
    1 & \text{otherwise}
    \end{cases}
\end{align*}
\]

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

\[ Y = X\beta + \sigma Z \]

First idea \( h_M(\beta_M) = I_{\{ |\beta_i| > \epsilon, i \in M \}} \) – issue: collinearity
Regression

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

\[
h_M(\beta_M) := I_{\left\{ \frac{1}{2} \| X^T (X_M \beta_M - X b_{\text{min}}) \|_2^2 \geq \epsilon(n, |M|) \right\}}
\]

where \( b_{\text{min}} \) solves

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

Regression

\[ Y = X\beta + \sigma Z \]

First idea: \( h_M(\beta_M) = I_{\{|\beta_i|>\epsilon, i\in M\}} \) – issue: collinearity

Better:

\[
h_M(\beta_M) := I\{\frac{1}{2} \|X^T(X_M\beta_M - Xb_{min})\|_2^2 \geq \epsilon(n, |M|)\}
\]

where \( b_{min} \) solves

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


similar to Dantzig selector Candes & Tao (2007)
different norm and target
GFD

\[ r(M|y) \propto \pi^{\frac{|M|}{2}} \Gamma\left(\frac{n - |M|}{2}\right) R^2 S_M^{-\left(\frac{n - |M| - 1}{2}\right)} E[h_M^\epsilon(\beta_M^*)] \]

Observations:

- Expectation with respect to within model GFD (usual T)
\[
 r(M|y) \propto \pi^{\frac{|M|}{2}} \Gamma\left(\frac{n - |M|}{2}\right) RSS_{M}^{-\frac{(n - |M| - 1)}{2}} E[h_{M}^{\epsilon}(\beta_{M}^{*})]
\]

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 \( r(M|y) = 0 \).
\[ r(M|y) \propto \pi^{\frac{|M|}{2}} \Gamma\left(\frac{n - |M|}{2}\right) \text{RSS}_M^{\left(\frac{n - |M| - 1}{2}\right)} E[h_M^\epsilon(\beta_M^*)] \]

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 \( r(M|y) = 0\).
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: $\liminf_{n \to \infty} \frac{n^{1-\alpha}}{\log(p)} > 2$, 

For the true model/parameter $p_T < \log n$

For a large model $j > p_T$ and large enough $n$ or $p$,

$\|X^T(\mathbf{H}_M \mathbf{H}_M(1))\|^2 < M(n; p)$,

where $\mathbf{H}_M$ and $\mathbf{H}_M(1)$ are the projection matrix for $M$ and $M$ with a covariate removed respectively.
Some Conditions

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

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

where \( b_{min} \) minimizes the norm subject to \( \| b \|_0 \leq p_T - 1 \).
Some Conditions

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

- For the true model/parameter \( p_T < \log n^\gamma \)

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

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 \( \rho = 0.6 \).
  - \( \varepsilon_M(n, p) = \Lambda_M \hat{\sigma}_M^2 \left( \frac{n^{0.51}}{9} + |M| \frac{\log(p\pi)^{1.1}}{9} - \log(n)^\gamma \right) \), 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?
Recall

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

$$\arg\min_{\xi} \{ \min_{\xi} \| x - G(U^*, \xi) \| \mid \| x - G(U^*, \xi) \| \leq \varepsilon \}$$
Recall

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

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

- Conditioning $U^*$ on $\{x = \mathbf{G}(U^*, \xi)\}$
  - “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 \text{) we get**
  \[ F^*(x_i^-) \leq U_i^* \leq F^*(x_i). \]
  There is a solution iff order of } U_i^* \text{ 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 \( x_i \).

- 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 $\mu^*, \sigma^*$ so that

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

![Graph showing a distribution with points marked at $x_i$, lower and upper limits, and a cumulative distribution function $F^*$]
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 $\mu^*, \sigma^*$ so that

  $$F(\sigma^{-1}(x_i - \mu^*)) = U_i^*,$$

  for all $i$.

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

\[0\] \[1\] 

\[N(4.5, 3^2)\] \[x_i\] lower upper
Constraint complications

Toy example: \( X = \mu + Z, \quad \mu > 0. \)
Constraint complications

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

- **Option 1:** condition \( Z^* \vert x - Z^* > 0 \)
  - \( r(\mu) = \frac{\varphi(x-\mu)}{\Phi(x)} I\{\mu > 0\} \)
  - Lower confidence bounds do not have correct coverage.
Constraint complications

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

- **Option 1:** condition \( Z^* | x - Z^* > 0 \)
  
  \[ r(\mu) = \frac{\varphi(x-\mu)}{\Phi(x)} I_{\{\mu > 0\}} \]
  
  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 \( Z^* | x - Z^* > 0 \)
  - \( r(\mu) = \frac{\varphi(x-\mu)}{\Phi(x)} I_{\{\mu > 0\}} \)
  - 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 \( \alpha \) 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 $U^*$ on concave solution (Gibbs sampler)
Example: Positive iid data with concave cdf (MLE is the Grenander estimator)

- Condition $U^*$ 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 $U^*$ 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 $U^*$ on concave solution (Gibbs sampler)
  Project unrestricted GFD to space of concave functions
  (quadratic program)
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

- Introduction
- Definition
- Sparsity
- Regularization
- Conclusions
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
What is it that we provide?

- GFI: General purpose method that often works well
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)
- Volatility estimation for high frequency data Katsoridas & H (2016+)
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!