Generalized Fiducial Inference

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

BFF 2018

Jan Hannig[a]

University of North Carolina at Chapel Hill

[a]NSF support acknowledged

Outline

- Introduction
- Definition
- Theoretical Results
- Applications
- Conclusions

Outline

- Introduction
- Definition
- Theoretical Results
- Applications
  - Distributed Data
  - Right Censored Data
  - High D Regression
- 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

Long, long, long time ago...

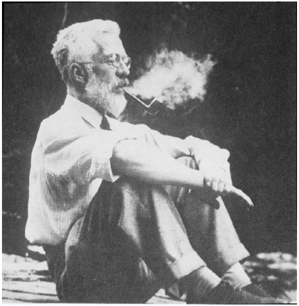
- Bayes Theorem: $P(\theta|X) = \frac{f(X|\theta)\pi(\theta)}{\int_{\Theta} f(X|\theta)\pi(\theta)d\theta}$.

Long, long, long time ago...

- Bayes Theorem: $P(\theta|X) = \frac{f(X|\theta)\pi(\theta)}{\int_\Theta f(X|\theta)\pi(\theta)d\theta}$.
- Bayes-Laplace postulate:

  When nothing is known about the parameter in advance, let the prior be so that all values of the parameter are equally likely.

Long, long, time ago...

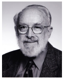


"Not knowing the chance of mutually exclusive events and knowing the chance to be equal are two quite different states of knowledge" R. A. Fisher (1930)

Brief history of fiducial inference

- Fisher (1922, 1930, 1935) no formal definition

Brief history of fiducial inference

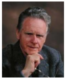


- Fisher (1922, 1930, 1935) no formal definition
- Lindley (1958) fiducial vs Bayes

Brief history of fiducial inference

- Fisher (1922, 1930, 1935) no formal definition
- Lindley (1958) fiducial vs Bayes
- Fraser (1966) structural inference

Brief history of fiducial inference



- Fisher (1922, 1930, 1935) no formal definition
- Lindley (1958) fiducial vs Bayes
- Fraser (1966) structural inference
- Dempster (1967) upper and lower probabilities

Brief history of fiducial inference

- Fisher (1922, 1930, 1935) no formal definition
- Lindley (1958) fiducial vs Bayes
- Fraser (1966) structural inference
- Dempster (1967) upper and lower probabilities
- Dawid and Stone (1982) theoretical results for simple cases.

Brief history of fiducial inference

- Fisher (1922, 1930, 1935) no formal definition
- Lindley (1958) fiducial vs Bayes
- Fraser (1966) structural inference
- Dempster (1967) upper and lower probabilities
- Dawid and Stone (1982) theoretical results for simple cases.
- Barnard (1995) pivotal based methods.

Brief history of fiducial inference

- Fisher (1922, 1930, 1935) no formal definition
- Lindley (1958) fiducial vs Bayes
- Fraser (1966) structural inference
- Dempster (1967) upper and lower probabilities
- Dawid and Stone (1982) theoretical results for simple cases.
- Barnard (1995) pivotal based methods.
- Weerahandi (1989, 1993) generalized inference.

Fiducial Inspired Inference since 2000

Fiducial Inspired Inference since 2000

- Dempster-Shafer calculus; Dempster (2008), Edlefsen, Liu & Dempster (2009)

Fiducial Inspired Inference since 2000

- Dempster-Shafer calculus; Dempster (2008), Edlefsen, Liu & Dempster (2009)
- Inferential Models; Liu & Martin (2015)

Fiducial Inspired Inference since 2000

- Dempster-Shafer calculus; Dempster (2008), Edlefsen, Liu & Dempster (2009)
- Inferential Models; Liu & Martin (2015)
- Confidence Distributions; Xie, Singh & Strawderman (2011), Schweder & Hjort (2016)

Fiducial Inspired Inference since 2000

- Dempster-Shafer calculus; Dempster (2008), Edlefsen, Liu & Dempster (2009)
- Inferential Models; Liu & Martin (2015)
- Confidence Distributions; Xie, Singh & Strawderman (2011), Schweder & Hjort (2016)
- Higher order likelihood, tangent exponential family, $r^\star$, Reid & Fraser (2010)

Fiducial Inspired Inference since 2000

- Dempster-Shafer calculus; Dempster (2008), Edlefsen, Liu & Dempster (2009)
- Inferential Models; Liu & Martin (2015)
- Confidence Distributions; Xie, Singh & Strawderman (2011), Schweder & Hjort (2016)
- Higher order likelihood, tangent exponential family, $r^\star$, Reid & Fraser (2010)
- Objective Bayesian inference, e.g., reference prior Berger, Bernardo & Sun (2009, 2012).

Fiducial Inspired Inference since 2000

- Dempster-Shafer calculus; Dempster (2008), Edlefsen, Liu & Dempster (2009)
- Inferential Models; Liu & Martin (2015)
- Confidence Distributions; Xie, Singh & Strawderman (2011), Schweder & Hjort (2016)
- Higher order likelihood, tangent exponential family, $r^\star$, Reid & Fraser (2010)
- Objective Bayesian inference, e.g., reference prior Berger, Bernardo & Sun (2009, 2012).
- Fiducial Inference H, Iyer & Patterson (2006), H (2009, 2013), H & Lee (2009), Taraldsen & Lindqvist (2013), Veronese & Melilli (2015), H, Iyer, Lai & Lee (2016)...

Aims

- Explain the definition of generalized fiducial distribution

Aims

- Explain the definition of generalized fiducial distribution
- Discuss theoretical results

Aims

- Explain the definition of generalized fiducial distribution
- Discuss theoretical results
- Show successful applications

Aims

- Explain the definition of generalized fiducial distribution
- Discuss theoretical results
- Show successful applications

- My point of view is frequentist
  - Justified using asymptotic theorems and simulations.
  - GFI shows very good repeated sampling performance in applications.

Outline

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?

Comparison to likelihood

- Data generating equation (DGE)

$$\mathbf{X} = \mathbf{G}(\mathbf{U}, \xi),$$

  - U is a random with known distribution (iid $U(0,1)$)
  - Parameter ξ is fixed.
  - Generate $\mathbf{X}$s by generating $\mathbf{U}$s and DGE.
    - This determines sampling distribution

Comparison to likelihood

- Data generating equation (DGE)

$$\mathbf{x} = \mathbf{G}(\mathbf{U}^{\star}, \xi),$$

  - U is a random with known distribution (iid $U(0,1)$)
  - Data $\mathbf{x}$ is fixed
  - Generate ξ by generating $\mathbf{U}$s and inverting DGE.
    - This determines fiducial distribution
    - Denote the inverse $Q_{\boldsymbol{x}}(\boldsymbol{U}^*)$.

Comparison to likelihood

- Data generating equation (DGE)

$$\mathbf{x} = \mathbf{G}(\mathbf{U}^{\star}, \xi),$$

  - U is a random with known distribution (iid $U(0,1)$)
  - Data $\mathbf{x}$ is fixed
  - Generate ξ by generating $\mathbf{U}$s and inverting DGE.
    - This determines fiducial distribution
    - Denote the inverse $Q_{\boldsymbol{x}}(\boldsymbol{U}^*)$.
- Issues: Multiple solutions and no solutions.

Example -- Bernoulli trials

- Data generating equation

$$X_i = 1\{U_i \leq p\},\ U_i \sim \text{Uniform}(0,1)$$

Generating U_i samples Bernoulli(p).

Example -- Bernoulli trials

- Data generating equation

$$X_i = 1\{U_i^\star \le p\},\ U_i^\star \sim \text{Uniform}(0,1)$$

Estimating U_i by $U_i^\star$ defines fiducial distribution

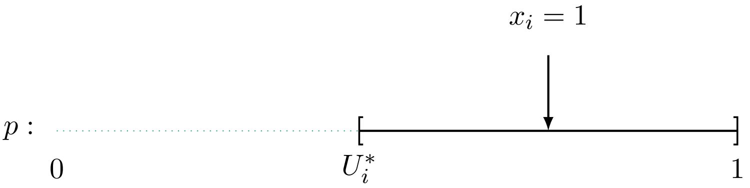
Example -- Bernoulli trials

- Data generating equation

$$X_i = 1\{U_i^\star \leq p\},\ U_i^\star \sim \text{Uniform}(0,1)$$

  Estimating U_i by $U_i^\star$ defines fiducial distribution

- If $x_i = 1$, then $p \in [U_i^\star, 1]$



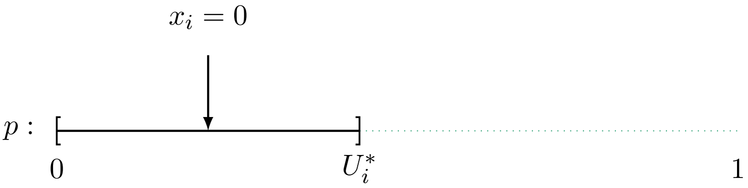
Example -- Bernoulli trials

- Data generating equation

$$X_i = 1\{U_i^\star \leq p\},\ U_i^\star \sim \text{Uniform}(0,1)$$

  Estimating U_i by $U_i^\star$ defines fiducial distribution

- If $x_i = 0$, then $p \in [0, U_i^\star]$



Example -- Binomial

- Data generating equation

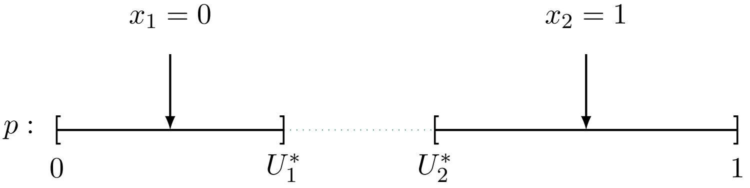
  $X_1 = 1\{U_1 \leq p\},\ X_2 = 1\{U_2 \leq p\}$ $\qquad$ U_1, U_2 i.i.d. Uniform(0,1)

Example -- Binomial

- Data generating equation

  $X_1 = 1\{U_1 \le p\},\ X_2 = 1\{U_2 \le p\}$ U_1, U_2 i.i.d. Uniform(0,1)

- If $X_1 = 0, X_2 = 1$ and $U_1^\star < U_2^\star$



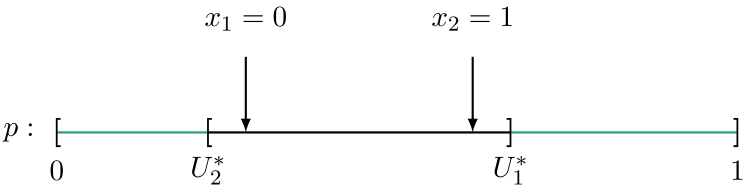
- No solution! Remove $(U_1^\star, U_2^\star)$ inconsistent with data.

Example -- Binomial

- Data generating equation

$$X_1 = 1\{U_1 \leq p\},\ X_2 = 1\{U_2 \leq p\} \qquad U_1, U_2 \text{ i.i.d. Uniform}(0,1)$$

- If $X_1 = 0, X_2 = 1$ and $U_1^\star > U_2^\star$



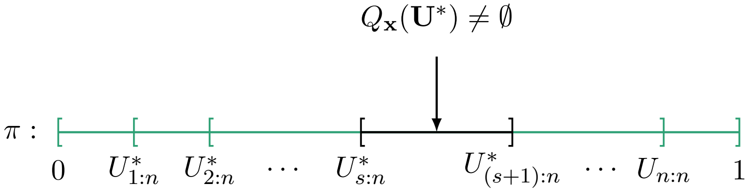
- $(U_1^\star, U_2^\star) \mid \{U_1^\star > U_2^\star\}$ estimates (u_1, u_2).

Example -- Binomial

- $(X_1, \dots X_n) \overset{iid}{\sim}$ Bernoulli(p), $S = \sum_{i=1}^{n} X_i \sim$ Binomial(n, p)

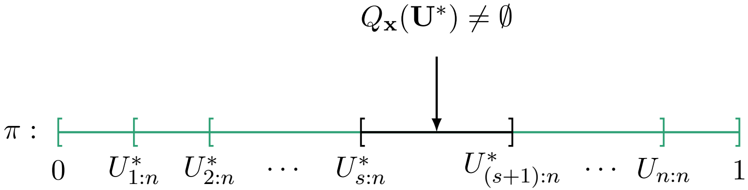
Example -- Binomial

- $(X_1, \dots X_n) \overset{iid}{\sim}$ Bernoulli(p), $S = \sum_{i=1}^{n} X_i \sim$ Binomial(n, p)
- Condition $\mathbf{U}^*$ on having a solution for p



Example -- Binomial

- $(X_1, \dots X_n) \overset{iid}{\sim}$ Bernoulli(p), $S = \sum_{i=1}^{n} X_i \sim$ Binomial(n, p)
- Condition $\mathbf{U}^*$ on having a solution for p



- May select a point in the interval; e.g. Beta$(s + 1/2, n - s + 1/2)$.

Example -- Location Cauchy

- Consider $X_i = \mu + U_i$ where U_i are i.i.d. standard Cauchy.

Example -- Location Cauchy

- Consider $X_i = \mu + U_i$ where U_i are i.i.d. standard Cauchy.
- Solve:

$$Q_{\boldsymbol{x}}(\boldsymbol{u}) = \begin{cases} x_1 - u_1 & \text{if } x_2 - x_1 = u_2 - u_1^\star, \ldots, x_n - x_1 = u_n - u_1 \\ \emptyset & \text{otherwise} \end{cases}$$

Example -- Location Cauchy

- Consider $X_i = \mu + U_i$ where U_i are i.i.d. standard Cauchy.
- Solve:

$$Q_{\boldsymbol{x}}(\boldsymbol{u}) = \begin{cases} x_1 - u_1 & \text{if } x_2 - x_1 = u_2 - u_1^\star, \ldots, x_n - x_1 = u_n - u_1 \\ \emptyset & \text{otherwise} \end{cases}$$

- Estimate $\boldsymbol{u}$ by conditional

$$\mathbf{U}^\star \mid x_2 - x_1 = U_2^\star - U_1^\star, \ldots, x_n - x_1 = U_n^\star - U_1^\star;$$

Example -- Location Cauchy

- Consider $X_i = \mu + U_i$ where U_i are i.i.d. standard Cauchy.
- Solve:

$$Q_{\boldsymbol{x}}(\boldsymbol{u}) = \begin{cases} x_1 - u_1 & \text{if } x_2 - x_1 = u_2 - u_1^\star, \ldots, x_n - x_1 = u_n - u_1 \\ \emptyset & \text{otherwise} \end{cases}$$

- Estimate $\boldsymbol{u}$ by conditional

$$\mathbf{U}^\star \mid x_2 - x_1 = U_2^\star - U_1^\star, \ldots, x_n - x_1 = U_n^\star - U_1^\star;$$

- Fiducial density $r(\mu|\boldsymbol{x}) \propto \prod_{i=1}^n (1 + (\mu - x_i)^2)^{-1}$.

Example -- Location Cauchy

- Consider $X_i = \mu + U_i$ where U_i are i.i.d. standard Cauchy.
- Solve:

$$Q_{\boldsymbol{x}}(\boldsymbol{u}) = \begin{cases} x_1 - u_1 & \text{if } x_2 - x_1 = u_2 - u_1^\star, \dots, x_n - x_1 = u_n - u_1 \\ \emptyset & \text{otherwise} \end{cases}$$

- Estimate $\boldsymbol{u}$ by conditional

$$\mathbf{U}^\star \mid x_2 - x_1 = U_2^\star - U_1^\star, \dots, x_n - x_1 = U_n^\star - U_1^\star;$$

- Fiducial density $r(\mu|\boldsymbol{x}) \propto \prod_{i=1}^n (1 + (\mu - x_i)^2)^{-1}$.
  - Location problem – same as posterior computed using Jeffreys prior

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

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

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

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

  - 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)

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

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

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

- $n = p$ gives $D(A) = |\det A|$
- $\|\bullet\|_2$ gives $D(A) = (\det A^\top A)^{1/2}$
- $\|\bullet\|_\infty$ gives $D(A) = \sum_{\mathbf{i}=(i_1,\ldots,i_p)} |\det(A)_{\mathbf{i}}|$

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 $X_i = \theta + (\theta^2 - \theta)U_i,\ U_i \sim U(0,1)$.
- $\frac{d}{d\theta}[\theta + (\theta^2 - \theta)U_i] = 1 + (2\theta - 1)U_i$, 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 $X_i = \theta + (\theta^2 - \theta)U_i,\ U_i \sim U(0,1)$.
- $\frac{d}{d\theta}[\theta + (\theta^2 - \theta)U_i] = 1 + (2\theta - 1)U_i$, with $U_i = \frac{X_i - \theta}{\theta^2 - \theta}$.
- Jacobian

$$J(\boldsymbol{x}, \theta) = D \begin{pmatrix} 1 + \frac{(2\theta-1)(x_1-\theta)}{\theta^2-\theta} \\ \vdots \\ 1 + \frac{(2\theta-1)(x_n-\theta)}{\theta^2-\theta} \end{pmatrix} = \frac{1}{\theta^2 - \theta} D \begin{pmatrix} x_1(2\theta - 1) - \theta^2 \\ \vdots \\ x_n(2\theta - 1) - \theta^2 \end{pmatrix}$$

Example -- Uniform(θ, θ^2)

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

$$J(\boldsymbol{x}, \theta) = D \begin{pmatrix} 1 + \frac{(2\theta-1)(x_1-\theta)}{\theta^2-\theta} \\ \vdots \\ 1 + \frac{(2\theta-1)(x_n-\theta)}{\theta^2-\theta} \end{pmatrix} = \frac{1}{\theta^2 - \theta} D \begin{pmatrix} x_1(2\theta - 1) - \theta^2 \\ \vdots \\ x_n(2\theta - 1) - \theta^2 \end{pmatrix}$$

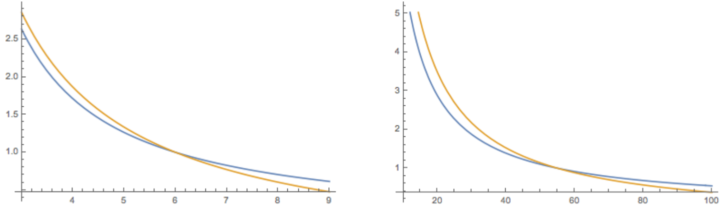
  - $= n\frac{\bar{x}(2\theta-1)-\theta^2}{\theta^2-\theta}$ for L_∞.

Example -- Uniform(θ, θ^2)

- Reference prior (Berger, Bernardo & Sun, 2009)
  $\pi(\theta) = \frac{e^{\psi\left(\frac{2\theta}{2\theta-1}\right)}(2\theta-1)}{\theta^2-\theta}$.

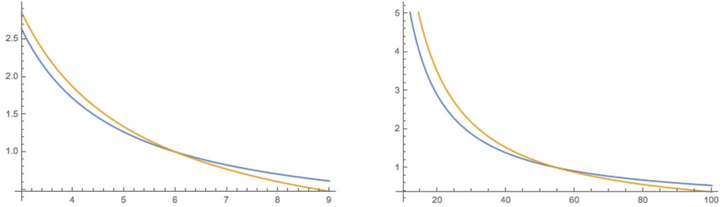
Example -- Uniform(θ, θ^2)

- Reference prior (Berger, Bernardo & Sun, 2009)
  $\pi(\theta) = \frac{e^{\psi\left(\frac{2\theta}{2\theta-1}\right)}(2\theta-1)}{\theta^2-\theta}$.
- reference prior vs fiducial Jacobian



Example -- Uniform(θ, θ^2)

- Reference prior (Berger, Bernardo & Sun, 2009)
  $\pi(\theta) = \frac{e^{\psi\left(\frac{2\theta}{2\theta-1}\right)}(2\theta-1)}{\theta^2-\theta}$.
- reference prior vs fiducial Jacobian



- In simulations fiducial was marginally better than reference prior which was much better than flat prior.

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 $Z = (Y - X\beta)/\sigma$.

Example -- Linear Regression

- Data generating equation $Y = X\beta + \sigma Z$
- $\frac{d}{d\theta}Y = (X, Z)$ and $Z = (Y - X\beta)/\sigma$.
- Jacobian $J(\boldsymbol{y}, \beta, \sigma) = D\left(\boldsymbol{X}, \frac{\boldsymbol{y}-\boldsymbol{X}\beta}{\sigma}\right) = \sigma^{-1}D(\boldsymbol{X}, \boldsymbol{y})$

Example -- Linear Regression

- Data generating equation $Y = X\beta + \sigma Z$
- $\frac{d}{d\theta} Y = (X, Z)$ and $Z = (Y - X\beta)/\sigma$.
- Jacobian $J(\boldsymbol{y}, \beta, \sigma) = D\left(\boldsymbol{X}, \frac{\boldsymbol{y} - \boldsymbol{X}\beta}{\sigma}\right) = \sigma^{-1} D(\boldsymbol{X}, \boldsymbol{y})$
  - $= \sigma^{-1} |\det(X^T X)|^{1/2} (RSS)^{1/2}$ for L_2.

Example -- Linear Regression

- Data generating equation $Y = X\beta + \sigma Z$
- $\frac{d}{d\theta} Y = (X, Z)$ and $Z = (Y - X\beta)/\sigma$.
- Jacobian $J(\boldsymbol{y}, \beta, \sigma) = D\left(\boldsymbol{X}, \frac{\boldsymbol{y}-\boldsymbol{X}\beta}{\sigma}\right) = \sigma^{-1} D(\boldsymbol{X}, \boldsymbol{y})$
  - $= \sigma^{-1} |\det(X^T X)|^{1/2} (RSS)^{1/2}$ for L_2.
- Same as independence Jeffreys, *explicit* normalizing constant

Example -- Generalized Pareto

- $X_i = G(U_i, \gamma, \sigma) = \sigma \frac{U_i^{-\gamma} - 1}{\gamma}$
  - Models excedances over a large threshold.

Example -- Generalized Pareto

- $X_i = G(U_i, \gamma, \sigma) = \sigma \frac{U_i^{-\gamma} - 1}{\gamma}$
  - Models excedances over a large threshold.
- Likelihood $f(\mathbf{x}, \gamma, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\left(1 + \frac{\gamma x_i}{\sigma}\right)^{1+1/\gamma}}$.

Example -- Generalized Pareto

- $X_i = G(U_i, \gamma, \sigma) = \sigma \frac{U_i^{-\gamma} - 1}{\gamma}$
  - Models excedances over a large threshold.
- Likelihood $f(\mathbf{x}, \gamma, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\left(1 + \frac{\gamma x_i}{\sigma}\right)^{1+1/\gamma}}$.
- Jacobian evaluated at $u_i = \left(1 + \frac{\gamma x_i}{\sigma}\right)^{-1/\gamma}$
  - $\frac{d}{d\sigma} G(u_i, \gamma, \sigma) = \frac{x_i}{\sigma}$.

Example -- Generalized Pareto

- $X_i = G(U_i, \gamma, \sigma) = \sigma \frac{U_i^{-\gamma} - 1}{\gamma}$
  - Models excedances over a large threshold.
- Likelihood $f(\mathbf{x}, \gamma, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\left(1+\frac{\gamma x_i}{\sigma}\right)^{1+1/\gamma}}$.
- Jacobian evaluated at $u_i = \left(1 + \frac{\gamma x_i}{\sigma}\right)^{-1/\gamma}$
  - $\frac{d}{d\sigma} G(u_i, \gamma, \sigma) = \frac{x_i}{\sigma}$.
  - $\frac{d}{d\gamma} G(u_i, \gamma, \sigma) = -\frac{x_i}{\gamma} + \frac{\sigma\left(1+\frac{\gamma x_i}{\sigma}\right)\log\left(1+\frac{\gamma x_i}{\sigma}\right)}{\gamma^2}$.

Example -- Generalized Pareto

- $X_i = G(U_i, \gamma, \sigma) = \sigma \frac{U_i^{-\gamma} - 1}{\gamma}$
  - Models excedances over a large threshold.
- Likelihood $f(\mathbf{x}, \gamma, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\left(1 + \frac{\gamma x_i}{\sigma}\right)^{1+1/\gamma}}$.
- Jacobian evaluated at $u_i = \left(1 + \frac{\gamma x_i}{\sigma}\right)^{-1/\gamma}$
  - $\frac{d}{d\sigma} G(u_i, \gamma, \sigma) = \frac{x_i}{\sigma}$.
  - $\frac{d}{d\gamma} G(u_i, \gamma, \sigma) = -\frac{x_i}{\gamma} + \frac{\sigma\left(1 + \frac{\gamma x_i}{\sigma}\right) \log\left(1 + \frac{\gamma x_i}{\sigma}\right)}{\gamma^2}$.
  - $J(\mathbf{x}, \gamma, \sigma) = \gamma^{-2} D \begin{pmatrix} x_1 & \left(1 + \frac{\gamma x_1}{\sigma}\right) \log\left(1 + \frac{\gamma x_1}{\sigma}\right) \\ \vdots & \vdots \\ x_n & \left(1 + \frac{\gamma x_n}{\sigma}\right) \log\left(1 + \frac{\gamma x_n}{\sigma}\right) \end{pmatrix}$

Example -- Generalized Pareto

- $X_i = G(U_i, \gamma, \sigma) = \sigma \frac{U_i^{-\gamma} - 1}{\gamma}$
  - Models excedances over a large threshold.
- Likelihood $f(\mathbf{x}, \gamma, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\left(1+\frac{\gamma x_i}{\sigma}\right)^{1+1/\gamma}}$.
- Jacobian evaluated at $u_i = \left(1 + \frac{\gamma x_i}{\sigma}\right)^{-1/\gamma}$
  - $\frac{d}{d\sigma} G(u_i, \gamma, \sigma) = \frac{x_i}{\sigma}$.
  - $\frac{d}{d\gamma} G(u_i, \gamma, \sigma) = -\frac{x_i}{\gamma} + \frac{\sigma\left(1+\frac{\gamma x_i}{\sigma}\right)\log\left(1+\frac{\gamma x_i}{\sigma}\right)}{\gamma^2}$.
  - $J(\mathbf{x}, \gamma, \sigma) = \gamma^{-2} D \begin{pmatrix} x_1 & \left(1 + \frac{\gamma x_1}{\sigma}\right) \log\left(1 + \frac{\gamma x_1}{\sigma}\right) \\ \vdots & \vdots \\ x_n & \left(1 + \frac{\gamma x_n}{\sigma}\right) \log\left(1 + \frac{\gamma x_n}{\sigma}\right) \end{pmatrix}$
    - $= \sum_{i<j} \left| \frac{x_j\left(1+\frac{\gamma x_i}{\sigma}\right)\log\left(1+\frac{\gamma x_i}{\sigma}\right) - x_i\left(1+\frac{\gamma x_j}{\sigma}\right)\log\left(1+\frac{\gamma x_j}{\sigma}\right)}{\gamma^2} \right|$ for L_∞.

Exercise

Derive a GFD for:

- Weibull distribution

Exercise

Derive a GFD for:

- Weibull distribution
- Negative Binomial Distribution (compare to Binomial)

Exercise

Derive a GFD for:

- Weibull distribution
- Negative Binomial Distribution (compare to Binomial)
- T distribution (might not have a pretty form)

Exercise

Derive a GFD for:

- Weibull distribution
- Negative Binomial Distribution (compare to Binomial)
- T distribution (might not have a pretty form)
- Your favorite model

Jacobian Formula

Short course special - L_1 norm!

Jacobian Formula

Short course special - L_1 norm!

- Recall:

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

Jacobian Formula

Short course special - L_1 norm!

- Recall:

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

- Optimization problem: $\min_{\xi} \sum_i |x_i - G_i(\mathbf{U}^{\star}, \xi)|$

Jacobian Formula

Short course special - L_1 norm!

- Recall:

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

- Optimization problem: $\min_{\xi} \sum_i |x_i - G_i(\mathbf{U}^{\star}, \xi)|$
  - $\boldsymbol{G}(\mathbf{U}^{\star}, \xi)$ is nearly linear in ξ on near $\mathbf{x}$

L_1 minimum

- Objective function: locally linear, locally convex

L_1 minimum

- Objective function: locally linear, locally convex
- At minimum:
  - p coordinates equal to $\mathbf{x}$
  - KKT condition:

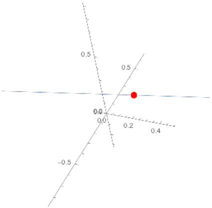
$$0 \in \partial \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^\star, \xi)\|_1,$$

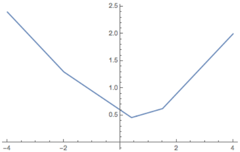
i.e., $0 = -\sum \lambda_i \nabla_x G_i(\mathbf{U}^\star, \xi)$, where

$$\lambda_i \in \begin{cases} \{1\} & x_i - G_i(\mathbf{U}^\star, \xi) > 0 \\ \{-1\} & x_i - G_i(\mathbf{U}^\star, \xi) < 0 \\ [-1, 1] & x_i - G_i(\mathbf{U}^\star, \xi) = 0 \end{cases}$$

$G = (.1, .25, -.2)^\top \xi + (.2, -.1, .3)^\top$

$\boldsymbol{x} = (0, 0, 0)^\top$





minimum at $\xi = .4,\ G = (0.24, 0, 0.22)$

L_1 Jacobian

- Select p equations $\boldsymbol{i} = (i_1, \ldots, i_p)$ and solve $\boldsymbol{x_i} = \boldsymbol{G_i}(U, \xi)$

L_1 Jacobian

- Select p equations $\boldsymbol{i} = (i_1, \ldots, i_p)$ and solve $\boldsymbol{x_i} = \boldsymbol{G_i}(U, \xi)$
  - Implicit function theorem: $\left| \det(\nabla_\xi \boldsymbol{G_i}(\boldsymbol{u}, \xi))_{\boldsymbol{u}=\mathbf{G}_{\boldsymbol{i}}^{-1}(\boldsymbol{x_i}, \xi)} \right| f(\boldsymbol{x_i}|\xi)$

L_1 Jacobian

- Select p equations $\boldsymbol{i} = (i_1, \ldots, i_p)$ and solve $\boldsymbol{x_i} = \boldsymbol{G_i}(U, \xi)$
  - Implicit function theorem: $\left|\det(\nabla_\xi \boldsymbol{G_i}(\boldsymbol{u}, \xi))_{\boldsymbol{u}=\mathbf{G}_{\boldsymbol{i}}^{-1}(\boldsymbol{x_i}, \xi)}\right| f(\boldsymbol{x_i} | \xi)$
- Condition on remaining of the equations and KKT condition

L_1 Jacobian

- Select p equations $\boldsymbol{i} = (i_1, \dots, i_p)$ and solve $\boldsymbol{x_i} = \boldsymbol{G_i}(U, \xi)$
  - Implicit function theorem: $\left|\det(\nabla_\xi \boldsymbol{G_i}(\boldsymbol{u}, \xi))_{\boldsymbol{u}=\mathbf{G}_{\boldsymbol{i}}^{-1}(\boldsymbol{x_i}, \xi)}\right| f(\boldsymbol{x_i}|\xi)$
- Condition on remaining of the equations and KKT condition
- Final formula $r(\xi|\boldsymbol{x}) \propto J(\boldsymbol{x}, \xi) f(\boldsymbol{x}|\xi)$,

$$J(\boldsymbol{x}, \xi) = \sum_{\boldsymbol{i}} k_{\boldsymbol{i}} \left|\det(\nabla_\xi \boldsymbol{G_i}(\boldsymbol{u}, \xi))_{\boldsymbol{u}=\mathbf{G}^{-1}(\boldsymbol{x}, \xi)}\right|.$$

L_1 Jacobian

- Select p equations $\boldsymbol{i} = (i_1, \ldots, i_p)$ and solve $\boldsymbol{x_i} = \boldsymbol{G_i}(U, \xi)$
  - Implicit function theorem: $\left|\det(\nabla_\xi \boldsymbol{G_i}(\boldsymbol{u}, \xi))_{\boldsymbol{u}=\mathbf{G}_{\boldsymbol{i}}^{-1}(\boldsymbol{x_i}, \xi)}\right| f(\boldsymbol{x_i}|\xi)$
- Condition on remaining of the equations and KKT condition
- Final formula $r(\xi|\boldsymbol{x}) \propto J(\boldsymbol{x}, \xi) f(\boldsymbol{x}|\xi)$,

$$J(\boldsymbol{x}, \xi) = \sum_{\boldsymbol{i}} k_{\boldsymbol{i}} \left| \det(\nabla_\xi \boldsymbol{G_i}(\boldsymbol{u}, \xi))_{\boldsymbol{u}=\mathbf{G}^{-1}(\boldsymbol{x}, \xi)} \right|.$$

  - The KKT factor

$$k_{\boldsymbol{i}} = P(\exists \lambda \in [-1, 1]^p \ : \ \lambda \cdot \nabla_\xi \boldsymbol{G_i}(\boldsymbol{u}, \xi) + R \cdot \nabla_\xi \boldsymbol{G_{-i}}(\boldsymbol{u}, \xi) = 0),$$

where $\boldsymbol{u} = \mathbf{G}^{-1}(\boldsymbol{x}, \xi)$,
$R = (R_1, \ldots, R_{n-p})$ i.i.d. Rademacher.

Outline

- Introduction
- Definition
- Theoretical Results
- Applications
  - Distributed Data
  - Right Censored Data
  - High D Regression
- Conclusions

Important Observations (Bayesian)

- GFD is always proper

Important Observations (Bayesian)

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

Important Observations (Bayesian)

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

Important Observations (Bayesian)

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

Important Observations (Bayesian)

- GFD is always proper
- GFD is invariant to re-parametrizations (same as Jeffreys)
- GFD is *not* invariant to smooth transformation of the data if $n > p$
- Consequently:
  - GFD does not satisfy likelihood principle.
  - Adding a multiple of a column to another column does not alter $D(A)$. Row operations not allowed!

Classical Result (n=1, p=1; Frequentist)

Data generating equation $S = G_S(\mathbf{U}, \xi)$ (1-dimensional statistic)

Classical Result (n=1, p=1; Frequentist)

Data generating equation $S = G_S(\mathbf{U}, \xi)$ (1-dimensional statistic)

1. $G_S(\mathbf{u}, \xi)$ is non-decreasing in ξ for all $\mathbf{u}$
2. For all $\mathbf{u}$ and s the inverse $Q_s(\mathbf{u}) = \{\xi : s = G_S(\mathbf{u}, \xi)\} \neq \emptyset$.
3. For all ξ the cdf $F_S(s, \xi)$ is continuous.

Classical Result (n=1, p=1; Frequentist)

Data generating equation $S = G_S(\mathbf{U}, \xi)$ (1-dimensional statistic)

1. $G_S(\mathbf{u}, \xi)$ is non-decreasing in ξ for all $\mathbf{u}$
2. For all $\mathbf{u}$ and s the inverse $Q_s(\mathbf{u}) = \{\xi : s = G_S(\mathbf{u}, \xi)\} \neq \emptyset$.
3. For all ξ the cdf $F_S(s, \xi)$ is continuous.

Then one has “unique” fiducial distribution and exact coverage of one-sided confidence intervals, i.e.,

$$P(Q_s(\mathbf{U}^\star) \leq \xi) = 1 - F_S(s, \xi).$$

Classical Result (n=1, p=1; Frequentist)

Data generating equation $S = G_S(\mathbf{U}, \xi)$ (1-dimensional statistic)

1. $G_S(\mathbf{u}, \xi)$ is non-decreasing in ξ for all $\mathbf{u}$
2. For all $\mathbf{u}$ and s the inverse $Q_s(\mathbf{u}) = \{\xi : s = G_S(\mathbf{u}, \xi)\} \neq \emptyset$.
3. For all ξ the cdf $F_S(s, \xi)$ is continuous.

Then one has "unique" fiducial distribution and exact coverage of one-sided confidence intervals, i.e.,

$$P(Q_s(\mathbf{U}^\star) \leq \xi) = 1 - F_S(s, \xi).$$

- If $S \sim \xi_0$ then $1 - F_S(S, \xi_0) \sim U(0, 1)$ – fiducial p-value.

Exact frequentist coverage

- Set $P(Q_s(\mathbf{U}^\star) \leq C_\alpha(s)) = 1 - \alpha$.

Exact frequentist coverage

- Set $P(Q_s(\mathbf{U}^\star) \leq C_\alpha(s)) = 1 - \alpha$.
- Coverage of upper confidence limit:

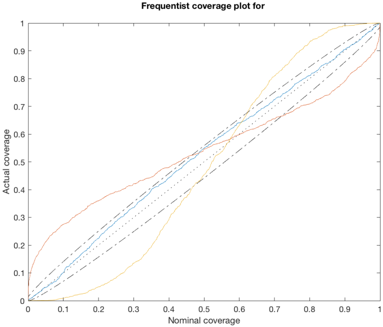
$$\begin{aligned} P_\xi(\xi \leq C_\alpha(S)) &= P_\xi(P(Q_S(\mathbf{U}^\star) \leq \xi | S) \leq 1 - \alpha) \\ &= P(U(0,1) \leq 1 - \alpha) = 1 - \alpha \end{aligned}$$

Exact frequentist coverage

- Set $P(Q_s(\mathbf{U}^\star) \le C_\alpha(s)) = 1 - \alpha$.
- Coverage of upper confidence limit:

$$P_\xi(\xi \le C_\alpha(S)) = P_\xi(P(Q_S(\mathbf{U}^\star) \le \xi | S) \le 1 - \alpha)$$
$$= P(U(0,1) \le 1 - \alpha) = 1 - \alpha$$

- This is general: simulate m fiducial p-values



exact
conservative
liberal

Exact frequentist coverage $(p > 1)$

- Which set of fiducial probability $1 - \alpha$ will be confidence set?

Exact frequentist coverage ($p > 1$)

- Which set of fiducial probability $1 - \alpha$ will be confidence set?
- Follow pivots and Inferential Models (Liu & Martin, 2015)

Exact frequentist coverage $(p > 1)$

- Which set of fiducial probability $1-\alpha$ will be confidence set?
- Follow pivots and Inferential Models (Liu & Martin, 2015)
  - Select a $P(\boldsymbol{U}^* \in \mathcal{U}) = \alpha$
  - Set $Q_S(\mathcal{U})$ has both fiducial probability and confidence $1-\alpha$

Exact frequentist coverage $(p > 1)$

- Which set of fiducial probability $1 - \alpha$ will be confidence set?
- Follow pivots and Inferential Models (Liu & Martin, 2015)
  - Select a $P(\boldsymbol{U}^* \in \mathcal{U}) = \alpha$
  - Set $Q_S(\mathcal{U})$ has both fiducial probability and confidence $1 - \alpha$
  - $p = 1$ above: $\xi \in (-\infty, C(s)) \longleftrightarrow u \in (\alpha, 1)$

Exact frequentist coverage $(p > 1)$

- Which set of fiducial probability $1-\alpha$ will be confidence set?
- Follow pivots and Inferential Models (Liu & Martin, 2015)
  - Select a $P(\boldsymbol{U}^* \in \mathcal{U}) = \alpha$
  - Set $Q_S(\mathcal{U})$ has both fiducial probability and confidence $1-\alpha$
  - $p = 1$ above: $\xi \in (-\infty, C(s)) \longleftrightarrow u \in (\alpha, 1)$
- Comments:
  - Links sets of $1-\alpha$ fiducial probability for different $\boldsymbol{X}$.

Exact frequentist coverage $(p > 1)$

- Which set of fiducial probability $1-\alpha$ will be confidence set?
- Follow pivots and Inferential Models (Liu & Martin, 2015)
  - Select a $P(\boldsymbol{U}^* \in \mathcal{U}) = \alpha$
  - Set $Q_S(\mathcal{U})$ has both fiducial probability and confidence $1-\alpha$
  - $p = 1$ above: $\xi \in (-\infty, C(s)) \longleftrightarrow u \in (\alpha, 1)$
- Comments:
  - Links sets of $1-\alpha$ fiducial probability for different $\boldsymbol{X}$.
  - Reverse: Map $C(S)$ of fiducial probability $1-\alpha$ to $\mathcal{U}$.
    If invariant in $\boldsymbol{X}$ then exact coverage.

Example -- Fieller's Problem

- $X \sim N(\mu_x, 1)$, $Y \sim N(\mu_y, 1)$, parameter of interest $\eta = \frac{\mu_x}{\mu_y}$

Example -- Fieller's Problem

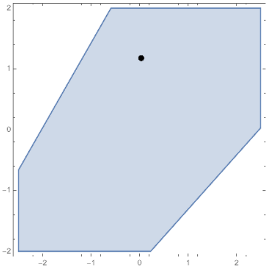
- $X \sim N(\mu_x, 1)$, $Y \sim N(\mu_y, 1)$, parameter of interest $\eta = \frac{\mu_x}{\mu_y}$
- DGE 1: $X = \mu_x + U_x$, $Y = \mu_y + U_y$
- Marginal Fiducial RV $Q_{x,y}(U_x^*, U_y^*) = \frac{x - U_x^*}{y - U_y^*}$

Example -- Fieller's Problem

- $X \sim N(\mu_x, 1),\ Y \sim N(\mu_y, 1)$, parameter of interest $\eta = \frac{\mu_x}{\mu_y}$
- DGE 1: $X = \mu_x + U_x,\ Y = \mu_y + U_y$
- Marginal Fiducial RV $Q_{x,y}(U_x^*, U_y^*) = \frac{x - U_x^*}{y - U_y^*}$
- Consider equal tailed regions of fiducial probability 95%:
  - When $|\mu_y| >> 0$ good frequentist performance, when $\mu_y \approx 0$ poor performance.

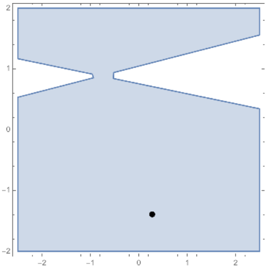
Example -- Fieller's Problem

- $X \sim N(\mu_x, 1)$, $Y \sim N(\mu_y, 1)$, parameter of interest $\eta = \frac{\mu_x}{\mu_y}$
- DGE 1: $X = \mu_x + U_x$, $Y = \mu_y + U_y$
- Marginal Fiducial RV $Q_{x,y}(U_x^*, U_y^*) = \frac{x - U_x^*}{y - U_y^*}$
- Consider equal tailed regions of fiducial probability 95%:
  - When $|\mu_y| >> 0$ good frequentist performance, when $\mu_y \approx 0$ poor performance.
  - Visualize $\mathcal{U} = \{(u_x, u_y) : \ c_1 \leq \frac{x - u_x^*}{y - u_y^*} \leq c_2\}$



$\mu_{x0} = 10, \mu_{y0} = 10$

coverage≈95%

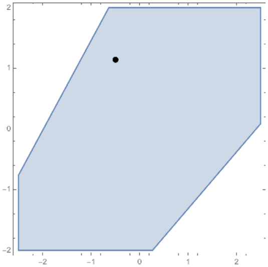


$\mu_{x0} = 0.5, \mu_{y0} = 0.5$

coverage≈98%

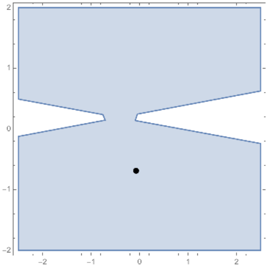
Example -- Fieller's Problem

- $X \sim N(\mu_x, 1)$, $Y \sim N(\mu_y, 1)$, parameter of interest $\eta = \frac{\mu_x}{\mu_y}$
- DGE 1: $X = \mu_x + U_x$, $Y = \mu_y + U_y$
- Marginal Fiducial RV $Q_{x,y}(U_x^*, U_y^*) = \frac{x-U_x^*}{y-U_y^*}$
- Consider equal tailed regions of fiducial probability 95%:
  - When $|\mu_y| >> 0$ good frequentist performance, when $\mu_y \approx 0$ poor performance.
  - Visualize $\mathcal{U} = \{(u_x, u_y) : \ c_1 \le \frac{x-u_x^*}{y-u_y^*} \le c_2\}$



$\mu_{x0} = 10, \mu_{y0} = 10$

coverage≈95%

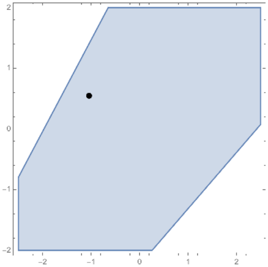


$\mu_{x0} = 0.5, \mu_{y0} = 0.5$

coverage≈98%

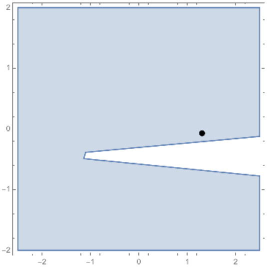
Example -- Fieller's Problem

- $X \sim N(\mu_x, 1)$, $Y \sim N(\mu_y, 1)$, parameter of interest $\eta = \frac{\mu_x}{\mu_y}$
- DGE 1: $X = \mu_x + U_x$, $Y = \mu_y + U_y$
- Marginal Fiducial RV $Q_{x,y}(U_x^*, U_y^*) = \frac{x - U_x^*}{y - U_y^*}$
- Consider equal tailed regions of fiducial probability 95%:
  - When $|\mu_y| >> 0$ good frequentist performance, when $\mu_y \approx 0$ poor performance.
  - Visualize $\mathcal{U} = \{(u_x, u_y) : \; c_1 \leq \frac{x - u_x^*}{y - u_y^*} \leq c_2\}$



$\mu_{x0} = 10, \mu_{y0} = 10$

coverage≈95%



$\mu_{x0} = 0.5, \mu_{y0} = 0.5$

coverage≈98%

Example -- Fieller's Solution

- DGE 2: $X = \eta\mu_y + \frac{-\eta U_1 + U_2}{\sqrt{1+\eta^2}}$, $Y = \mu_y + \frac{U_1 + \eta U_2}{\sqrt{1+\eta^2}}$

Example -- Fieller's Solution

- DGE 2: $X = \eta\mu_y + \frac{-\eta U_1 + U_2}{\sqrt{1+\eta^2}}$, $Y = \mu_y + \frac{U_1 + \eta U_2}{\sqrt{1+\eta^2}}$
- Fiducial density $r(\eta|x,y) = \frac{e^{-\frac{(x-\eta y)^2}{2(\eta^2+1)}}|y+x\eta|}{\sqrt{2\pi}(\eta^2+1)^{3/2}}$
  - Jacobian $J(x,y,\eta,\mu_y) = \frac{|y+x\eta|}{1+\eta^2}$

Example -- Fieller's Solution

- DGE 2: $X = \eta\mu_y + \frac{-\eta U_1 + U_2}{\sqrt{1+\eta^2}}$, $Y = \mu_y + \frac{U_1 + \eta U_2}{\sqrt{1+\eta^2}}$
- Fiducial density $r(\eta|x,y) = \frac{e^{-\frac{(x-\eta y)^2}{2(\eta^2+1)}}|y+x\eta|}{\sqrt{2\pi}(\eta^2+1)^{3/2}}$
  - Jacobian $J(x,y,\eta,\mu_y) = \frac{|y+x\eta|}{1+\eta^2}$
- Fieller picked the set of fiducial probability 95% corresponding to $\mathcal{U} = \{|u_1| < 1.96\}$

Example -- Fieller's Solution

- DGE 2: $X = \eta\mu_y + \frac{-\eta U_1 + U_2}{\sqrt{1+\eta^2}}$, $Y = \mu_y + \frac{U_1 + \eta U_2}{\sqrt{1+\eta^2}}$
- Fiducial density $r(\eta|x,y) = \frac{e^{-\frac{(x-\eta y)^2}{2(\eta^2+1)}} |y+x\eta|}{\sqrt{2\pi}(\eta^2+1)^{3/2}}$
  - Jacobian $J(x,y,\eta,\mu_y) = \frac{|y+x\eta|}{1+\eta^2}$
- Fieller picked the set of fiducial probability 95% corresponding to $\mathcal{U} = \{|u_1| < 1.96\}$
  - Pros: Fiducial sets are linked, exact coverage is guaranteed
  - Cons: The shape of the sets is strange (interval, complement of interval, whole real line)

Example -- Fieller's Solution

- DGE 2: $X = \eta\mu_y + \frac{-\eta U_1 + U_2}{\sqrt{1+\eta^2}}$, $Y = \mu_y + \frac{U_1 + \eta U_2}{\sqrt{1+\eta^2}}$
- Fiducial density $r(\eta|x,y) = \frac{e^{-\frac{(x-\eta y)^2}{2(\eta^2+1)}}|y+x\eta|}{\sqrt{2\pi}(\eta^2+1)^{3/2}}$
  - Jacobian $J(x,y,\eta,\mu_y) = \frac{|y+x\eta|}{1+\eta^2}$
- Fieller picked the set of fiducial probability 95% corresponding to $\mathcal{U} = \{|u_1| < 1.96\}$
  - Pros: Fiducial sets are linked, exact coverage is guaranteed
  - Cons: The shape of the sets is strange (interval, complement of interval, whole real line)
- GFD1 $\approx$ GFD2 if $|y| >> 0$.

Ancillary Representation $(n > 1, p = 1)$

(4) Let $(S(\mathbf{X}), \mathbf{A}(\mathbf{X}))$ be a smooth 1-1 transformation of $\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{U}, \xi)$.

- $S(\mathbf{X})$ is one dimensional satisfying 1, 2, 3.
- $\mathbf{A}(\mathbf{X})$ is a vector of functional ancillary statistics $(\frac{\partial}{\partial \xi}\mathbf{A} \circ \mathbf{G}(\boldsymbol{U}, \xi) = \mathbf{0})$.

Theorem (Majumder, 2015)

If (4) is satisfied GFI derived from $(S, \mathbf{A})$ is exact.

Ancillary Representation $(n > 1, p = 1)$

(4) Let $(S(\mathbf{X}), \mathbf{A}(\mathbf{X}))$ be a smooth 1-1 transformation of $\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{U}, \xi)$.

- $S(\mathbf{X})$ is one dimensional satisfying 1, 2, 3.
- $\mathbf{A}(\mathbf{X})$ is a vector of functional ancillary statistics $(\frac{\partial}{\partial \xi}\mathbf{A} \circ \mathbf{G}(\boldsymbol{U}, \xi) = \mathbf{0})$.

Theorem (Majumder, 2015)

If (4) is satisfied GFI derived from $(S, \mathbf{A})$ is exact.

- Idea: The GFD is the same as FD based on $S = G_{S|\mathbf{a}}(\mathbf{U}_\mathbf{a}, \xi)$
  - $\mathbf{U}_\mathbf{a} \sim \mathbf{U} \mid \mathbf{a} = A(\mathbf{G}(\mathbf{U}, \xi))$
  - $G_{S|\mathbf{a}}$ is the restriction of G to the domain of $U_\mathbf{a}$.

Ancillary Representation $(n > 1, p = 1)$

(4) Let $(S(\mathbf{X}), \mathbf{A}(\mathbf{X}))$ be a smooth 1-1 transformation of $\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{U}, \xi)$.
- $S(\mathbf{X})$ is one dimensional satisfying 1, 2, 3.
- $\mathbf{A}(\mathbf{X})$ is a vector of functional ancillary statistics $(\frac{\partial}{\partial \xi}\mathbf{A} \circ \mathbf{G}(\boldsymbol{U}, \xi) = \mathbf{0})$.

Theorem (Majumder, 2015)

If (4) is satisfied GFI derived from $(S, \mathbf{A})$ is exact.

- Idea: The GFD is the same as FD based on $S = G_{S|\mathbf{a}}(\mathbf{U_a}, \xi)$
  - $\mathbf{U_a} \sim \mathbf{U} \mid \mathbf{a} = A(\mathbf{G}(\mathbf{U}, \xi))$
  - $G_{S|\mathbf{a}}$ is the restriction of G to the domain of $U_{\mathbf{a}}$.
- Same argument works for $p > 1$.

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+).

Another look

- Do fiducial probabilities correspond to (asymptotic) coverage?
- What is needed?

Another look

- Do fiducial probabilities correspond to (asymptotic) coverage?
- What is needed?
  - Convergence of the posteriors to something nice

Another look

- Do fiducial probabilities correspond to (asymptotic) coverage?
- What is needed?
  - Convergence of the posteriors to something nice
  - Linkage of credible sets across all potential data.

LARGE

Data $\boldsymbol{X}_n$ generated using $\xi_{n,0}$ and GFD measures $R_{n,\boldsymbol{X}_n}$.

LARGE

Data $\boldsymbol{X}_n$ generated using $\xi_{n,0}$ and GFD measures $R_{n,\boldsymbol{X}_n}$.

Bernstein-von Mises Thm: centered and scaled MLE

1. $t_n(\boldsymbol{X}_n) \xrightarrow{\mathcal{D}} \boldsymbol{T} = (\boldsymbol{T}_1, \boldsymbol{T}_2)$
2. $\boldsymbol{T}_1 = \boldsymbol{H}_1(\boldsymbol{V}_1, \boldsymbol{\zeta}), \boldsymbol{T}_2 = \boldsymbol{H}_2(\boldsymbol{V}_2)$
   - $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are one-to-one
     $Q_{\boldsymbol{t}_1}(\boldsymbol{v}_1) = \zeta$ solves $\boldsymbol{t}_1 = \boldsymbol{H}_1(\boldsymbol{v}_1, \boldsymbol{\zeta})$
     GFD $R_{\boldsymbol{t}} \sim Q_{\boldsymbol{t}_1}(\boldsymbol{V}_1^\star) \mid \boldsymbol{H}_2(\boldsymbol{V}_2^\star) = \boldsymbol{t}_2$.
   - $R_{\boldsymbol{t}}(\partial C) = 0$ for open C
3. Homeomorphic injective $\boldsymbol{\Xi}_n$
   - $\Xi_n(\boldsymbol{\zeta}_{n,0}) = 0$
   - $t_n(\boldsymbol{x}_n) \to \boldsymbol{t}$

   implies $R_{n,\boldsymbol{x}_n}\boldsymbol{\Xi}_n^{-1} \xrightarrow{\mathcal{W}} R_{\boldsymbol{t}}$.

$\boldsymbol{T} = \boldsymbol{\zeta} + \boldsymbol{V},\ \boldsymbol{V} \sim N(0, I_{\boldsymbol{\xi}_0}^{-1})$

$N(\boldsymbol{t}, I_{\boldsymbol{\xi}_0}^{-1})$

$\boldsymbol{\Xi}_n(\boldsymbol{\xi}) = \sqrt{n}(\boldsymbol{\xi} - \boldsymbol{\xi}_0)$

LARGE theorem

Theorem (H., Lai & Lee, 2016)

Assume LARGE, fix $0 < \alpha < 1$, select open $C_n(\boldsymbol{x}_n)$

1. $R_{n,\boldsymbol{x}_n}(C_n(\boldsymbol{x}_n)) = \alpha$
2. $t_n(\boldsymbol{x}_n) \to \boldsymbol{t}$ *implies* $\boldsymbol{\Xi}_n(C_n(\boldsymbol{x}_n)) \to C(\boldsymbol{t})$
3. *the set* $\mathcal{V}_{\boldsymbol{t}_2} = \{\boldsymbol{v} : \boldsymbol{t} = \boldsymbol{H}(\boldsymbol{v}, \xi)$ *for* $\xi \in C(\boldsymbol{t})\}$ *depends on* $\boldsymbol{t} = (\boldsymbol{t}_1, \boldsymbol{t}_2)$ *only through ancillary* $\boldsymbol{t}_2$.

LARGE theorem

Theorem (H., Lai & Lee, 2016)

Assume LARGE, fix $0 < \alpha < 1$, select open $C_n(\boldsymbol{x}_n)$

1. $R_{n,\boldsymbol{x}_n}(C_n(\boldsymbol{x}_n)) = \alpha$
2. $t_n(\boldsymbol{x}_n) \to \boldsymbol{t}$ *implies* $\boldsymbol{\Xi}_n(C_n(\boldsymbol{x}_n)) \to C(\boldsymbol{t})$
3. *the set* $\mathcal{V}_{\boldsymbol{t}_2} = \{\boldsymbol{v} : \boldsymbol{t} = \boldsymbol{H}(\boldsymbol{v}, \xi) \text{ for } \xi \in C(\boldsymbol{t})\}$ *depends on* $\boldsymbol{t} = (\boldsymbol{t}_1, \boldsymbol{t}_2)$ *only through ancillary* $\boldsymbol{t}_2$.

Then the sets $C_n(\boldsymbol{x}_n)$ are α asymptotic confidence sets.

LARGE theorem

Theorem (H., Lai & Lee, 2016)

Assume LARGE, fix $0 < \alpha < 1$, select open $C_n(\boldsymbol{x}_n)$

1. $R_{n,\boldsymbol{x}_n}(C_n(\boldsymbol{x}_n)) = \alpha$
2. $t_n(\boldsymbol{x}_n) \to \boldsymbol{t}$ *implies* $\boldsymbol{\Xi}_n(C_n(\boldsymbol{x}_n)) \to C(\boldsymbol{t})$
3. *the set* $\mathcal{V}_{\boldsymbol{t}_2} = \{\boldsymbol{v} : \boldsymbol{t} = \boldsymbol{H}(\boldsymbol{v}, \xi)$ *for* $\xi \in C(\boldsymbol{t})\}$ *depends on* $\boldsymbol{t} = (\boldsymbol{t}_1, \boldsymbol{t}_2)$ *only through ancillary* $\boldsymbol{t}_2$*.*

Then the sets $C_n(\boldsymbol{x}_n)$ are α asymptotic confidence sets.

- If $\boldsymbol{T}_1 = \boldsymbol{H}_1(\boldsymbol{V}_1, \boldsymbol{\xi})$ has group structure – 3) means $C(\boldsymbol{t})$ is invariant in $\boldsymbol{t}_1$.

Example - Uniform(θ, θ^2)

The conditions are satisfied

Example - Uniform(θ, θ^2)

The conditions are satisfied

- Fiducial Density $r(\theta|x) \propto \frac{\bar{x}(2\theta-1)-\theta^2}{(\theta^2-\theta)^{n+1}} I_{(\sqrt{x_{(n)}}, x_{(1)})}(\theta)$

Example - Uniform(θ, θ^2)

The conditions are satisfied

- Fiducial Density $r(\theta|x) \propto \frac{\bar{x}(2\theta-1)-\theta^2}{(\theta^2-\theta)^{n+1}} I_{(\sqrt{x_{(n)}}, x_{(1)})}(\theta)$
- Transformations:

$$t_n(\boldsymbol{y}) = n \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix} \cdot \begin{pmatrix} x_{(1)} - \theta_0 \\ \frac{\theta_0^2 - x_{(n)}}{2\theta_0} \end{pmatrix}, \quad \Xi_n(\theta) = n(\theta - \theta_0)$$

Example - Uniform(θ, θ^2)

The conditions are satisfied

- Fiducial Density $r(\theta|x) \propto \frac{\bar{x}(2\theta-1)-\theta^2}{(\theta^2-\theta)^{n+1}} I_{(\sqrt{x_{(n)}}, x_{(1)})}(\theta)$
- Transformations:
  $t_n(\boldsymbol{y}) = n \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix} \cdot \begin{pmatrix} x_{(1)} - \theta_0 \\ \frac{\theta_0^2 - x_{(n)}}{2\theta_0} \end{pmatrix}, \quad \Xi_n(\theta) = n(\theta - \theta_0)$
- Limit $T_1 = \xi + V_1,\ T_2 = V_2,$ where V_1, V_2 are dependent.
- Limiting fiducial distribution
  $r(\zeta|\boldsymbol{t}) \propto \exp\left(-\zeta \frac{2\theta_0 - 1}{\theta_0^2 - \theta_0}\right) I_{(T_1 - T_2, T_1 + T_2)}(\zeta).$

Example - Uniform(θ, θ^2)

The conditions are satisfied

- Fiducial Density $r(\theta|x) \propto \frac{\bar{x}(2\theta-1)-\theta^2}{(\theta^2-\theta)^{n+1}} I_{(\sqrt{x_{(n)}},x_{(1)})}(\theta)$
- Transformations:
$$t_n(\boldsymbol{y}) = n \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix} \cdot \begin{pmatrix} x_{(1)} - \theta_0 \\ \frac{\theta_0^2 - x_{(n)}}{2\theta_0} \end{pmatrix}, \quad \Xi_n(\theta) = n(\theta - \theta_0)$$
- Limit $T_1 = \xi + V_1,\ T_2 = V_2,$ where V_1, V_2 are dependent.
- Limiting fiducial distribution
$$r(\zeta|\boldsymbol{t}) \propto \exp\left(-\zeta \frac{2\theta_0-1}{\theta_0^2-\theta_0}\right) I_{(T_1-T_2,T_1+T_2)}(\zeta).$$
- One sided credible sets have correct asymptotic coverage!

Outline

Computational Issues

- When lucky: Direct Monte Carlo inversion of DGE

Computational Issues

- When lucky: Direct Monte Carlo inversion of DGE
- Typically need a normalizing constant to $J(\boldsymbol{x}, \xi) f(\boldsymbol{x}|\xi)$
  - Ride the Bayesian Wave:
    MCMC (Gibbs, MH), SMC (trick is to resample right)

Computational Issues

- When lucky: Direct Monte Carlo inversion of DGE
- Typically need a normalizing constant to $J(\boldsymbol{x},\xi)f(\boldsymbol{x}|\xi)$
  - Ride the Bayesian Wave:
    MCMC (Gibbs, MH), SMC (trick is to resample right)
  - For p small – numerical integration

Computational Issues

- When lucky: Direct Monte Carlo inversion of DGE
- Typically need a normalizing constant to $J(\boldsymbol{x}, \xi) f(\boldsymbol{x}|\xi)$
  - Ride the Bayesian Wave:
    MCMC (Gibbs, MH), SMC (trick is to resample right)
  - For p small – numerical integration

- Fiducial Quirk – Interested in Frequentist Properties
  - m-number of MC samples vs k-number of replications:
    $k \sim Cm^2$

Outline

- Introduction
- Definition
- Theoretical Results
- Applications
  - Distributed Data
  - Right Censored Data
  - High D Regression
- Conclusions

Distributed Data

- Motivation
  - n is so big that the $\boldsymbol{X}$'s cannot be loaded to one computer

Distributed Data

- Motivation
  - n is so big that the $\boldsymbol{X}$'s cannot be loaded to one computer
  - the data are at different sites

Distributed Data

- Motivation
  - n is so big that the $\boldsymbol{X}$'s cannot be loaded to one computer
  - the data are at different sites
  - data cannot be released off site for privacy concerns

Distributed Data

- Motivation
  - n is so big that the $\boldsymbol{X}$'s cannot be loaded to one computer
  - the data are at different sites
  - data cannot be released off site for privacy concerns
- partition $\boldsymbol{x}$ into K subsets, where each subsets can be analyzed

Distributed Data

- Motivation
  - n is so big that the $\boldsymbol{X}$'s cannot be loaded to one computer
  - the data are at different sites
  - data cannot be released off site for privacy concerns
- partition $\boldsymbol{x}$ into K subsets, where each subsets can be analyzed
  - e.g., use a computer cluster for parallel processing, where K is the number of nodes (or workers)

Distributed Data

- Motivation
  - n is so big that the $\boldsymbol{X}$'s cannot be loaded to one computer
  - the data are at different sites
  - data cannot be released off site for privacy concerns
- partition $\boldsymbol{x}$ into K subsets, where each subsets can be analyzed
  - e.g., use a computer cluster for parallel processing, where K is the number of nodes (or workers)
  - merge results from different nodes

Importance sampling for Massive Data

Importance sampling for Massive Data

- $\boldsymbol{x} = \boldsymbol{x}_1 \cup \boldsymbol{x}_2 \cup \ldots \cup \boldsymbol{x}_K$

Importance sampling for Massive Data

- $\boldsymbol{x} = \boldsymbol{x}_1 \cup \boldsymbol{x}_2 \cup \ldots \cup \boldsymbol{x}_K$
- $r(\boldsymbol{\theta}|\boldsymbol{x})$ – the generalized fiducial density of $\boldsymbol{x}$
- $r(\boldsymbol{\theta}|\boldsymbol{x}_k)$ – the generalized fiducial density of $\boldsymbol{x}_k$

Importance sampling for Massive Data

- $\boldsymbol{x} = \boldsymbol{x}_1 \cup \boldsymbol{x}_2 \cup \ldots \cup \boldsymbol{x}_K$
- $r(\boldsymbol{\theta}|\boldsymbol{x})$ – the generalized fiducial density of $\boldsymbol{x}$
- $r(\boldsymbol{\theta}|\boldsymbol{x}_k)$ – the generalized fiducial density of $\boldsymbol{x}_k$
- On each worker sample from $q_k(\boldsymbol{\theta})$

$$r(\boldsymbol{\theta}|\boldsymbol{x}) \propto \sum_k \text{'importance weight'} \times q_k(\boldsymbol{\theta})$$

Importance sampling for Massive Data

- $\boldsymbol{x} = \boldsymbol{x}_1 \cup \boldsymbol{x}_2 \cup \ldots \cup \boldsymbol{x}_K$
- $r(\boldsymbol{\theta}|\boldsymbol{x})$ – the generalized fiducial density of $\boldsymbol{x}$
- $r(\boldsymbol{\theta}|\boldsymbol{x}_k)$ – the generalized fiducial density of $\boldsymbol{x}_k$
- On each worker sample from $q_k(\boldsymbol{\theta})$

$$r(\boldsymbol{\theta}|\boldsymbol{x}) \propto \sum_k \text{'importance weight'} \times q_k(\boldsymbol{\theta})$$

- depending on the problem, $r(\boldsymbol{\theta}|\boldsymbol{x})$ and $r_k(\boldsymbol{\theta}|\boldsymbol{x}_k)$ may be known only up to normalizing constant.

Naive scheme

- generate a fiducial sample for data on each node

Naive scheme

- generate a fiducial sample for data on each node
- compute the weight $w_k(\boldsymbol{\theta}) = \frac{r(\boldsymbol{\theta}|\boldsymbol{x})}{r_k(\boldsymbol{\theta}|\boldsymbol{x}_k)}$

Naive scheme

- generate a fiducial sample for data on each node
- compute the weight $w_k(\boldsymbol{\theta}) = \frac{r(\boldsymbol{\theta}|\boldsymbol{x})}{r_k(\boldsymbol{\theta}|\boldsymbol{x}_k)}$
- Not feasible and very inefficient!

Naive scheme

- generate a fiducial sample for data on each node
- compute the weight $w_k(\boldsymbol{\theta}) = \frac{r(\boldsymbol{\theta}|\boldsymbol{x})}{r_k(\boldsymbol{\theta}|\boldsymbol{x}_k)}$
- Not feasible and very inefficient!
  - Target of order $n^{-1/2}$
  - Fiducial sample on each worker of order $n_k^{-1/2}$.
  - Most realizations get extremely small weights.

Improved scheme

Improved scheme

- Each worker computes MLE $\hat{\theta}_k$ and empirical Fisher Information $\hat{I}_k$ and passes it to other workers

Improved scheme

- Each worker computes MLE $\hat{\theta}_k$ and empirical Fisher Information $\hat{I}_k$ and passes it to other workers
- Each worker simulates a sample from $q(\mathbf{x}_k) \propto r_k(\boldsymbol{\theta}|\boldsymbol{x}_k) \times \prod_{j \neq k} g(\boldsymbol{\theta}|\hat{\theta}_j, \hat{I}_j)$

Improved scheme

- Each worker computes MLE $\hat{\theta}_k$ and empirical Fisher Information $\hat{I}_k$ and passes it to other workers
- Each worker simulates a sample from $q(\mathbf{x}_k) \propto r_k(\boldsymbol{\theta}|\boldsymbol{x}_k) \times \prod_{j\neq k} g(\boldsymbol{\theta}|\hat{\theta}_j, \hat{I}_j)$
  - Practical choice $g \sim \text{Normal}(\hat{\theta}_j, \gamma \hat{I}_k^{-1})$.

Improved scheme

- Each worker computes MLE $\hat{\theta}_k$ and empirical Fisher Information $\hat{I}_k$ and passes it to other workers
- Each worker simulates a sample from $q(\mathbf{x}_k) \propto r_k(\boldsymbol{\theta}|\boldsymbol{x}_k) \times \prod_{j \neq k} g(\boldsymbol{\theta}|\hat{\theta}_j, \hat{I}_j)$
  - Practical choice $g \sim$ Normal$(\hat{\theta}_j, \gamma \hat{I}_k^{-1})$.
- Weight $w_k(\boldsymbol{\theta}) \approx \prod_{j \neq k} \frac{f(\mathbf{x}_k, \boldsymbol{\theta})}{g(\boldsymbol{\theta}|\theta_j, I_j)}$
  (Computed in parallel.)

Improved scheme

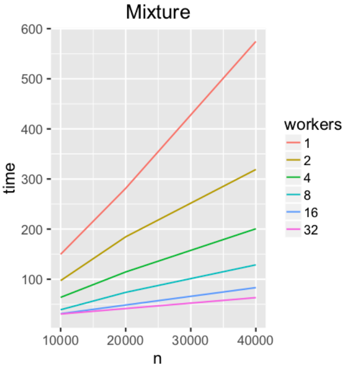
- Each worker computes MLE $\hat{\theta}_k$ and empirical Fisher Information $\hat{I}_k$ and passes it to other workers
- Each worker simulates a sample from $q(\mathbf{x}_k) \propto r_k(\boldsymbol{\theta}|\boldsymbol{x}_k) \times \prod_{j \neq k} g(\boldsymbol{\theta}|\hat{\theta}_j, \hat{I}_j)$
  - Practical choice $g \sim$ Normal$(\hat{\theta}_j, \gamma \hat{I}_k^{-1})$.
- Weight $w_k(\boldsymbol{\theta}) \approx \prod_{j \neq k} \frac{f(\mathbf{x}_k, \boldsymbol{\theta})}{g(\boldsymbol{\theta}|\theta_j, I_j)}$
  (Computed in parallel.)

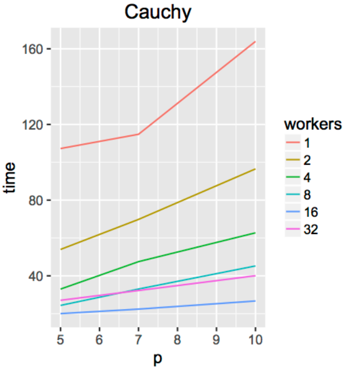
- We proved consistency and asymptotic normality of the approximation error.

Experiments

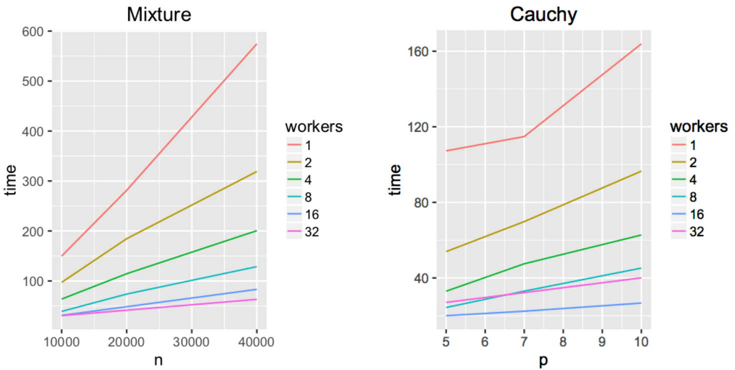
- All good performance ($K = 1, 2, 4, 8, 16, 32$)
  - Gaussian mixture: $0.6N(-1, 1) + 0.4N(1, 1)$ ($n = 4 \times 10^5$)
  - Linear regression with Cauchy errors ($n = 10^5$, $p_T = 4, p = 6, 8, 11$)

Computational time



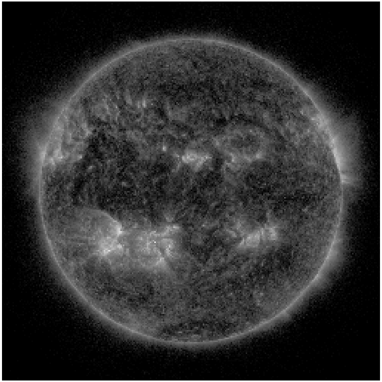


Computational time

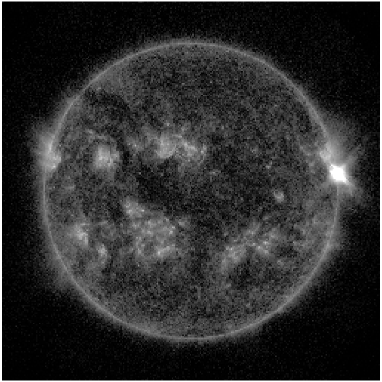


- For Cauchy: speed improves until $K = 16$ then deteriorates (Cheng & Shang, 2015)

Sun Spots

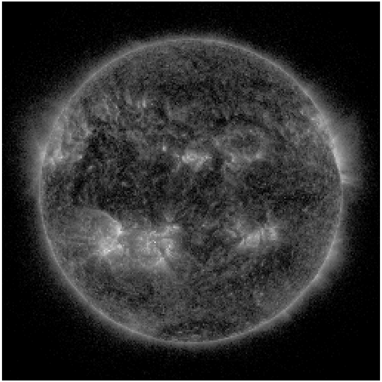


low activity

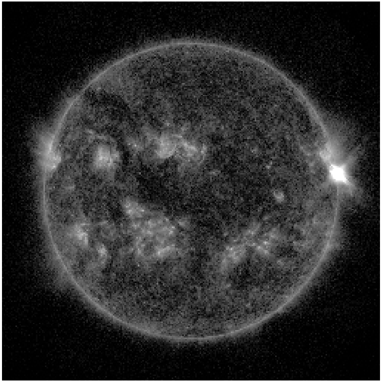


high activity

Sun Spots



low activity



high activity

- The bright flare on the right has value 253. Is this high?

Data

- Solar Dynamics Observatory (SDO), launched on 2010

Data

- Solar Dynamics Observatory (SDO), launched on 2010
- one instrument is Atmospheric Imaging Assembly (AIA) (Schuh et al. 2013)
  - photographs the sun in 8 wavelengths every 12s
  - image size: 4096×4096
  - 1.5 TB compressed data per day
  - same as 3 TB raw (i.e., uncompressed) data per day

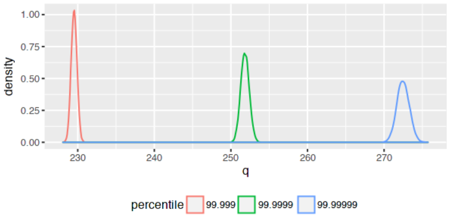
Data

- Solar Dynamics Observatory (SDO), launched on 2010
- one instrument is Atmospheric Imaging Assembly (AIA) (Schuh et al. 2013)
  - photographs the sun in 8 wavelengths every 12s
  - image size: 4096×4096
  - 1.5 TB compressed data per day
  - same as 3 TB raw (i.e., uncompressed) data per day
- ultimate goal: detect and predict solar flares

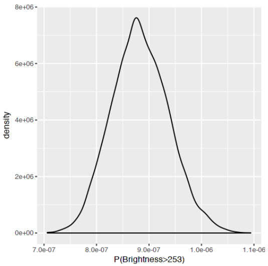
Data

- Solar Dynamics Observatory (SDO), launched on 2010
- one instrument is Atmospheric Imaging Assembly (AIA) (Schuh et al. 2013)
  - photographs the sun in 8 wavelengths every 12s
  - image size: 4096×4096
  - 1.5 TB compressed data per day
  - same as 3 TB raw (i.e., uncompressed) data per day
- ultimate goal: detect and predict solar flares
- Tool: GFD for Generalized Pareto (Wandler & H, 2012)

GFD for extreme quantiles



Large Quantiles



Fiducial Excedance Probability

Outline

Right Censored Data

- Data generating equation:

  $T_i = F^{-1}(U_i),\ C_i = G_i^{-1}(W_i|F^{-1}(U_i)), \quad U_i, W_i$ i.i.d. Uniform(0,1)

  observe only $X_i = \min(T_i, C_i), \delta_i = I_{\{T_i \leq C_i\}}$.

Right Censored Data

- Data generating equation:

  $T_i = F^{-1}(U_i),\ C_i = G_i^{-1}(W_i|F^{-1}(U_i)), \quad U_i, W_i$ i.i.d. Uniform(0,1)

  observe only $X_i = \min(T_i, C_i), \delta_i = I_{\{T_i \le C_i\}}$.
- Solvie for F and G:

$$\begin{array}{lll} \delta_i = 1 & F^*(x_i - \epsilon) < U_i^\star \le F^*(x_i) & G_i^*(x_i - \epsilon|x_i) \le W_i^\star \\ \delta_i = 0 & F^*(x_i) < U_i^\star & G_i^*(x_i - \epsilon|F^{-1}(U_i^\star)) < W_i^\star \le G_i^*(x_i|F^{-1}(U_i^\star)) \end{array}$$

Right Censored Data

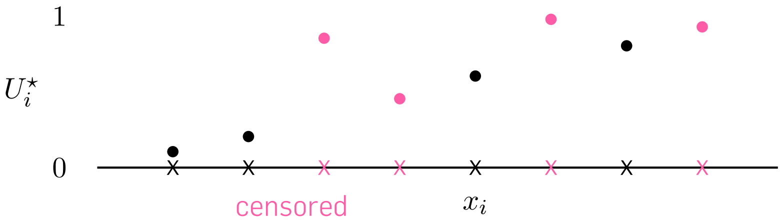
- Data generating equation:

  $$T_i = F^{-1}(U_i),\ C_i = G_i^{-1}(W_i|F^{-1}(U_i)), \quad U_i, W_i \text{ i.i.d. Uniform}(0,1)$$

  observe only $X_i = \min(T_i, C_i), \delta_i = I_{\{T_i \le C_i\}}$.
- Solvie for F and G:

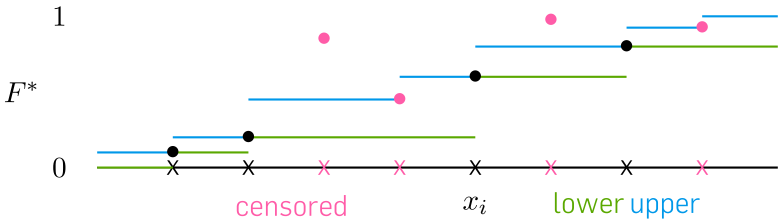
  $$\begin{array}{lll} \delta_i = 1 & F^*(x_i - \epsilon) < U_i^\star \le F^*(x_i) & G_i^*(x_i - \epsilon | x_i) \le W_i^\star \\ \delta_i = 0 & F^*(x_i) < U_i^\star & G_i^*(x_i - \epsilon | F^{-1}(U_i^\star)) < W_i^\star \le G_i^*(x_i | F^{-1}(U_i^\star)) \end{array}$$
- Generate U^* conditional on existence of F^* solving the equations.

Visual Demonstation



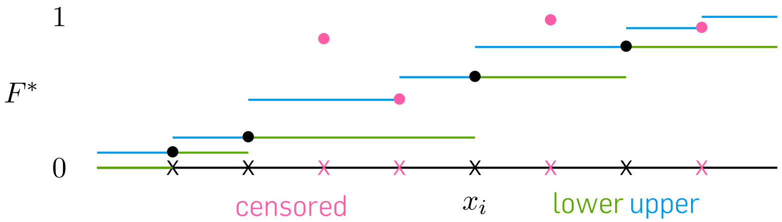
U_i^* generated so that there is a solution

Visual Demonstation



F^* is any cdf between bounds

Visual Demonstation



F^* is any cdf between bounds

Fact: For any failure time $E^*[F_L^\star(s)] < \hat{F}(s) < E^*[F_U^\star(s)]$.

Theoretical Results

- Known: KM estimator

$$\sqrt{n}[\hat{F}(t) - F(t)] \xrightarrow{\mathcal{D}} W_t \text{ in D[0,T]}$$

(W – mean zero Gaussian process with known covariance depending on censoring)

Theoretical Results

- Known: KM estimator

$$\sqrt{n}[\hat{F}(t) - F(t)] \xrightarrow{\mathcal{D}} W_t \text{ in D[0,T]}$$

(W – mean zero Gaussian process with known covariance depending on censoring)

- Theorem (Cui, Hannig): Under similar assumptions

$$\sqrt{n}[F^{\star}(t) - \hat{F}(t)]|\mathbf{X} \xrightarrow{\mathcal{D}} W_t \text{ almost surely.}$$

Implementation

- Fast MC algorithm for generating samples from fiducial distribution.

Implementation

- Fast MC algorithm for generating samples from fiducial distribution.
- Quantiles of fiducial at given x – approximate pointwise CIs.

Implementation

- Fast MC algorithm for generating samples from fiducial distribution.
- Quantiles of fiducial at given x – approximate pointwise CIs.
- Simultaneous CI
  - Center on mean (or Kaplan Mayer).
  - Find L_∞ ball containing $1-\alpha$ fiducial sample curves.

Implementation

- Fast MC algorithm for generating samples from fiducial distribution.
- Quantiles of fiducial at given x – approximate pointwise CIs.
- Simultaneous CI
  - Center on mean (or Kaplan Mayer).
  - Find L_∞ ball containing $1-\alpha$ fiducial sample curves.
- Fiducial p-value for testing $H_0 : F = F_0$
  - Find largest α so that F_0 is in the $1-\alpha$ CI.

Implementation

- Fast MC algorithm for generating samples from fiducial distribution.
- Quantiles of fiducial at given x – approximate pointwise CIs.
- Simultaneous CI
  - Center on mean (or Kaplan Mayer).
  - Find L_∞ ball containing $1-\alpha$ fiducial sample curves.
- Fiducial p-value for testing $H_0 : F = F_0$
  - Find largest α so that F_0 is in the $1-\alpha$ CI.
- For difference of two populations – use difference of fiducial distributions

Lessons from Simulation

We simulated several one and two sample settings

Lessons from Simulation

We simulated several one and two sample settings

- Pointwise simulations show good coverage

Lessons from Simulation

We simulated several one and two sample settings

- Pointwise simulations show good coverage
- Simultaneous CI also show good/conservative coverage

Lessons from Simulation

We simulated several one and two sample settings

- Pointwise simulations show good coverage
- Simultaneous CI also show good/conservative coverage
- Two sample testing
  - Good type 1 error
  - Compared to log rank tests and sup-log rank test: competitive.

Lessons from Simulation

We simulated several one and two sample settings

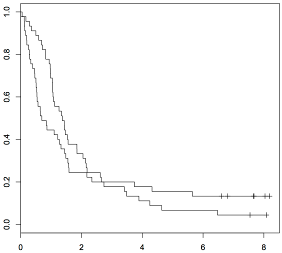
- Pointwise simulations show good coverage
- Simultaneous CI also show good/conservative coverage
- Two sample testing
  - Good type 1 error
  - Compared to log rank tests and sup-log rank test: competitive.
- Room for improvement:
  - Use something else than L_∞ ball – half-region depth?

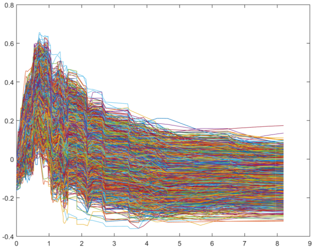
Gastric Tumor Study Group (1982)

- Clinical trial of chemotherapy vs. chemotherapy combined with radiotherapy.

Gastric Tumor Study Group (1982)

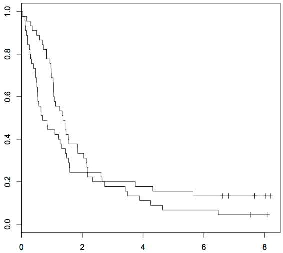
- Clinical trial of chemotherapy vs. chemotherapy combined with radiotherapy.

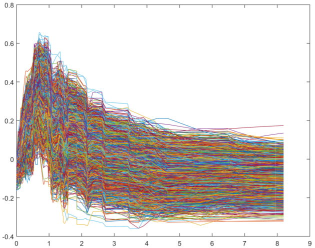




Gastric Tumor Study Group (1982)

- Clinical trial of chemotherapy vs. chemotherapy combined with radiotherapy.

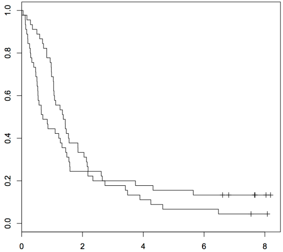


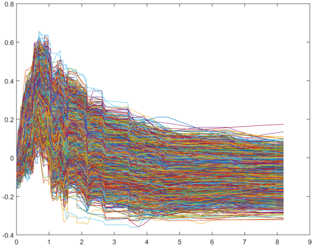


- Different: Log-rank statistics p-value 0.63 (hazard crossing); Fiducial p-value 0.003.

Gastric Tumor Study Group (1982)

- Clinical trial of chemotherapy vs. chemotherapy combined with radiotherapy.





- Different: Log-rank statistics p-value 0.63 (hazard crossing); Fiducial p-value 0.003.
- Simulation with estimated hazard shows fiducial more powerful than competitors.

Outline

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 - see poster!

- Penalty is used to discourage models with many parameters

Alternative to penalty - see poster!

- 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 - see poster!

- 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

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

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

Alternative to penalty - see poster!

- 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

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

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

- 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:

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

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:

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

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

Regression

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

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

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
  - Call this: ε-admissible subset

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 $p_T < \log n^\gamma$

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

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

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 $p_T < \log n^\gamma$

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

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,

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

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

Default ε

$$\varepsilon = \Lambda_M \widehat{\sigma}^2_M \left(\frac{n^{0.51}}{9} + |M| \frac{\log(p\pi)^{1.1}}{9} - p_T \right)_+,$$

- $\Lambda_M := \operatorname{tr}\left((H_M X)' H_M X\right)$ with $H_M := X_M (X_M' X_M)^{-1} X_M'$
- $\widehat{\sigma}^2_M := \mathsf{RSS}_M / (n - |M|)$

Default ε

$$\varepsilon = \Lambda_M \widehat{\sigma}^2_M \left(\frac{n^{0.51}}{9} + |M| \frac{\log(p\pi)^{1.1}}{9} - p_T \right)_+ ,$$

- $\Lambda_M := \operatorname{tr}\big((H_M X)' H_M X\big)$ with $H_M := X_M (X_M' X_M)^{-1} X_M'$
- $\widehat{\sigma}^2_M :=$ $\mathsf{RSS}_M/(n - |M|)$
- Tuning parameter p_T represents belief about true $|M_T|$.

Simulation setup 1

- Generate 1000 data vectors y from linear model with $\beta^0_{M_o} = (-1.5, -1, -.8, -.6, .6, .8, 1, 1.5)'$, and $\sigma^0_{M_o} = 1$.

Simulation setup 1

- Generate 1000 data vectors y from linear model with $\beta^0_{M_o} = (-1.5, -1, -.8, -.6, .6, .8, 1, 1.5)'$, and $\sigma^0_{M_o} = 1$.
- The $n \times p$ design matrix X is generated with rows from the $\mathsf{N}_p(0, \Sigma)$ distribution, where the diagonal components $\Sigma_{ii} = 1$ and the off-diagonal components $\Sigma_{ij} = \rho$ for $i \neq j$.

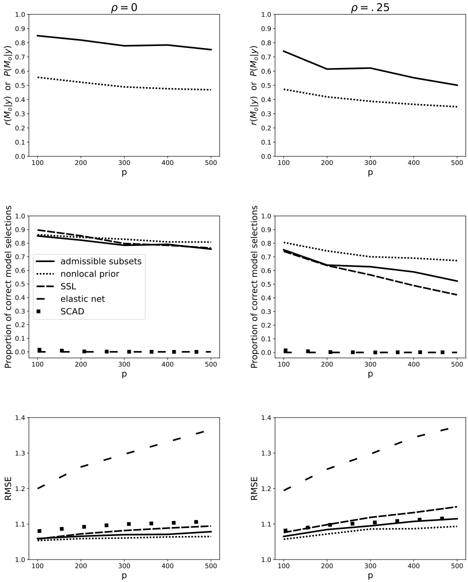
Simulation setup 1

- Generate 1000 data vectors y from linear model with $\beta^0_{M_o} = (-1.5, -1, -.8, -.6, .6, .8, 1, 1.5)'$, and $\sigma^0_{M_o} = 1$.
- The $n \times p$ design matrix X is generated with rows from the $\mathsf{N}_p(0, \Sigma)$ distribution, where the diagonal components $\Sigma_{ii} = 1$ and the off-diagonal components $\Sigma_{ij} = \rho$ for $i \neq j$.
- Implement 10-fold cross-validation scheme for choosing the tuning parameter p_o (prior to starting the algorithm).

Simulation setup 1

- Generate 1000 data vectors y from linear model with $\beta^0_{M_o} = (-1.5, -1, -.8, -.6, .6, .8, 1, 1.5)'$, and $\sigma^0_{M_o} = 1$.
- The $n \times p$ design matrix X is generated with rows from the $\mathrm{N}_p(0, \Sigma)$ distribution, where the diagonal components $\Sigma_{ii} = 1$ and the off-diagonal components $\Sigma_{ij} = \rho$ for $i \neq j$.
- Implement 10-fold cross-validation scheme for choosing the tuning parameter p_o (prior to starting the algorithm).
- Set $n = 100$, and consider $p = 100, 200, 300, 400, 500$.

Simulation results 1



Simulation setup 2

To illustrate the difference from the nonlocal prior approach, for $n = 30$, generate data from the following model.

$$Y \sim \mathsf{N}_n\Big(1 \cdot x^{(1)} + 1 \cdot x^{(2)} + \cdots + 1 \cdot x^{(9)}, I_n\Big),$$

where $x^{(1)}, x^{(2)}, x^{(3)} \overset{\text{iid}}{\sim} \mathsf{N}_n(0, I_n)$, and

$$\begin{array}{ccccccccc}
x^{(4)} & \sim & \mathsf{N}_n\Big(& .25 \cdot x^{(1)} & & & & & , .1^2 I_n\Big) \\
x^{(5)} & \sim & \mathsf{N}_n\Big(& & & .5 \cdot x^{(2)} & & & , .1^2 I_n\Big) \\
x^{(6)} & \sim & \mathsf{N}_n\Big(& & & & - & .75 \cdot x^{(3)} & , .1^2 I_n\Big) \\
x^{(7)} & \sim & \mathsf{N}_n\Big(& x^{(1)} & & & + & x^{(3)} & , .1^2 I_n\Big) \\
x^{(8)} & \sim & \mathsf{N}_n\Big(& & & x^{(2)} & - & x^{(3)} & , .1^2 I_n\Big) \\
x^{(9)} & \sim & \mathsf{N}_n\Big(& x^{(1)} & + & x^{(2)} & + & x^{(3)} & , .1^2 I_n\Big)
\end{array}$$

Simulation results 2

	MAP size	RMSE	$P(M_{\mathsf{MAP}}\|y)$
ε-$admissible$ subsets	3.476	1.138	.365
nonlocal prior	8.997	1.197	.333

- RMSE of an out-of-sample test set of 30 observations
- Averaged over 1000 synthetic data sets

Simulation results 2

	MAP size	RMSE	$P(M_{\text{MAP}}\|y)$
ε-$admissible$ subsets	3.476	1.138	.365
nonlocal prior	8.997	1.197	.333

- RMSE of an out-of-sample test set of 30 observations
- Averaged over 1000 synthetic data sets
- Nonlocal prior procedure typically includes all 9 covariates even though the y can be mostly explained by 3.

Outline

BFF

- Many great minds contributed to foundations of statistics in the past – Fisher, Neyman, de Finetti, Lindley, Savage, LeCam, Cox, Efron, Berger, Fraser, Reid, Dempster, Dawid, ...

BFF

- Many great minds contributed to foundations of statistics in the past – Fisher, Neyman, de Finetti, Lindley, Savage, LeCam, Cox, Efron, Berger, Fraser, Reid, Dempster, Dawid, ...
  - Area was not known for harmonious relationships and respectful discourse

the "protracted battle" among leading statistics founding fathers "has left statistics without a philosophy that matches contemporary attitudes." (Kass, 2011)

BFF

- Many great minds contributed to foundations of statistics in the past – Fisher, Neyman, de Finetti, Lindley, Savage, LeCam, Cox, Efron, Berger, Fraser, Reid, Dempster, Dawid, ...
  - Area was not known for harmonious relationships and respectful discourse

the "protracted battle" among leading statistics founding fathers "has left statistics without a philosophy that matches contemporary attitudes." (Kass, 2011)

Can Bayesian, Fiducial and Frequentist
become Best Friends Forever?

Fiducial Future

Fiducial Future

- What is it that we provide? “Can we solve something others cannot?”
  - GFI: General purpose method that often works well

Fiducial Future

- What is it that we provide? “Can we solve something others cannot?”
  - GFI: General purpose method that often works well
- Computational convenience and efficiency
  - Fiducial options in software
  - Deep learning?

Fiducial Future

- What is it that we provide? “Can we solve something others cannot?”
  - GFI: General purpose method that often works well
- Computational convenience and efficiency
  - Fiducial options in software
  - Deep learning?
- New kind of theoretical guarantees

Fiducial Future

- What is it that we provide? “Can we solve something others cannot?”
  - GFI: General purpose method that often works well
- Computational convenience and efficiency
  - Fiducial options in software
  - Deep learning?
- New kind of theoretical guarantees
- Applications
  - The proof is in the pudding

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 people will gain 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 people will gain similar trust in fiducial inspired approaches.

Thank you!