Statistical Inference based on Inverse Data Generating Equation

(Generalized Fiducial Inference)

Jan Hannig

jan.hannig@unc.edu

University of North Carolina at Chapel Hill

Hari Iyer, Colorado State University; Derek Sonderegger, University of Northern Arizona;

Jessi Cisewski, Carnegie Mellon University; Thomas C.M. Lee, Randy Lai, University of California at Davis;

Fiducial?

- Fiducial inference was mentioned only briefly during my graduate studies. I did not remember what it was about. The only think that stuck in my mind was that it is “bad”.

Fiducial?

- Fiducial inference was mentioned only briefly during my graduate studies. I did not remember what it was about. The only think that stuck in my mind was that it is “bad”.
- 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

Brief history of fiducial inference

- Fisher (1930) introduced the idea of fiducial probability and inference in an attempt to overcome what he saw as a serious deficiency of the Bayesian approach to inference – use of a prior distribution when no prior information was available.

$$r(\xi|x) = -\frac{\partial F(x|\xi)}{\partial \xi}.$$

Brief history of fiducial inference

- Fisher (1930) introduced the idea of fiducial probability and inference in an attempt to overcome what he saw as a serious deficiency of the Bayesian approach to inference – use of a prior distribution when no prior information was available.

$$r(\xi|x) = -\frac{\partial F(x|\xi)}{\partial \xi}.$$

- Fisher (1935) further elaborated on this idea. E.g., to eliminate nuisance parameters he suggested substituting their fiducial distribution. As an example he considered the inference for the difference of two normal means – "Behrens-Fisher problem" .

Brief history of fiducial inference

- Fraser (1960) – structural inference

Brief history of fiducial inference

- Fraser (1960) – structural inference
- Dawid and Stone (1982) theoretical results for simple cases.

Brief history of fiducial inference

- Fraser (1960) – structural inference
- Dawid and Stone (1982) theoretical results for simple cases.
- Bernard (1995) pivotal based methods.

Brief history of fiducial inference

- Fraser (1960) – structural inference
- Dawid and Stone (1982) theoretical results for simple cases.
- Bernard (1995) pivotal based methods.
- Weerahandi (1989, 1993) generalized inference.

Brief history of fiducial inference

- Fraser (1960) – structural inference
- Dawid and Stone (1982) theoretical results for simple cases.
- Bernard (1995) pivotal based methods.
- Weerahandi (1989, 1993) generalized inference.
- Hannig, Iyer, Patterson (2006) generalized inference is closely related to fiducial inference & theoretical properties.

Related Modern Research

- Dempster-Shafer calculus; Dempster (2008). Similar to fiducial inference. Designed for discrete distributions; leads to upper and lower probabilities.

Related Modern Research

- Dempster-Shafer calculus; Dempster (2008). Similar to fiducial inference. Designed for discrete distributions; leads to upper and lower probabilities.
- Inferential Models; Liu, Martin and coworkers.

Related Modern Research

- Dempster-Shafer calculus; Dempster (2008). Similar to fiducial inference. Designed for discrete distributions; leads to upper and lower probabilities.
- Inferential Models; Liu, Martin and coworkers.
- Confidence Distributions; Xie, Singh & Strawderman (2011), Schweder & Hjort (2002) The idea is to use a frequentist procedure (e.g., one sided CI for all possible confidence levels α) to define a distribution on the parameter space.

Related Modern Research

- Dempster-Shafer calculus; Dempster (2008). Similar to fiducial inference. Designed for discrete distributions; leads to upper and lower probabilities.
- Inferential Models; Liu, Martin and coworkers.
- Confidence Distributions; Xie, Singh & Strawderman (2011), Schweder & Hjort (2002) The idea is to use a frequentist procedure (e.g., one sided CI for all possible confidence levels α) to define a distribution on the parameter space.
- Objective Bayesian inference; choice of $\pi(\theta)$ when we have no prior info, e.g., reference prior Berger, Bernardo & Sun (2009).

Related Modern Research

- Dempster-Shafer calculus; Dempster (2008). Similar to fiducial inference. Designed for discrete distributions; leads to upper and lower probabilities.
- Inferential Models; Liu, Martin and coworkers.
- Confidence Distributions; Xie, Singh & Strawderman (2011), Schweder & Hjort (2002) The idea is to use a frequentist procedure (e.g., one sided CI for all possible confidence levels α) to define a distribution on the parameter space.
- Objective Bayesian inference; choice of $\pi(\theta)$ when we have no prior info, e.g., reference prior Berger, Bernardo & Sun (2009).
  - With improper reference prior one needs to prove that the posterior is a proper distribution on an individual basis.

Bird's Eye of Statistical Inference

- We are given a data set X and are asked to provide some information about the mechanism used to generate it.

Bird's Eye of Statistical Inference

- We are given a data set X and are asked to provide some information about the mechanism used to generate it.
- Frequentist Inference
  - Assume that the data was generated using a model $\mathcal{P} = \{P_\theta\}_{\theta \in \times}$.

Bird's Eye of Statistical Inference

- We are given a data set X and are asked to provide some information about the mechanism used to generate it.
- Frequentist Inference
  - Assume that the data was generated using a model $\mathcal{P} = \{P_\theta\}_{\theta \in \times}$.
  - Goal is to find a P_θ's that best fit the data with possible some additional considerations, e.g., sparsity.

Bird's Eye of Statistical Inference

- We are given a data set X and are asked to provide some information about the mechanism used to generate it.
- Frequentist Inference
  - Assume that the data was generated using a model $\mathcal{P} = \{P_\theta\}_{\theta \in \times}$.
  - Goal is to find a P_θ's that best fit the data with possible some additional considerations, e.g., sparsity.
  - Each statistical problem requires its own solution and the quality of the solution is judged by repeated sampling performance (Cournot's principle).

Bird's Eye of Statistical Inference

- Bayesian inference
  - It is assumed that the value theta $\theta \in \Theta$ was generated using some known distribution $\pi(\theta)$, prior, and we have only single, fully known distribution $P_\theta \cdot \pi(\theta)$.

Bird's Eye of Statistical Inference

- Bayesian inference
  - It is assumed that the value theta $\theta \in \Theta$ was generated using some known distribution $\pi(\theta)$, prior, and we have only single, fully known distribution $P_\theta \cdot \pi(\theta)$.
  - The random variable θ is unobserved and needs to be predicted, using standard statistical techniques – Bayes theorem.

Bird's Eye of Statistical Inference

- Bayesian inference
  - It is assumed that the value theta $\theta \in \Theta$ was generated using some known distribution $\pi(\theta)$, prior, and we have only single, fully known distribution $P_\theta \cdot \pi(\theta)$.
  - The random variable θ is unobserved and needs to be predicted, using standard statistical techniques – Bayes theorem. The predictive distribution $\theta|\boldsymbol{X}$, posterior, has subjective interpretation (de Finetti's betting interpretation).

Bird's Eye of Statistical Inference

- Bayesian inference
  - It is assumed that the value theta $\theta \in \Theta$ was generated using some known distribution $\pi(\theta)$, prior, and we have only single, fully known distribution $P_\theta \cdot \pi(\theta)$.
  - The random variable θ is unobserved and needs to be predicted, using standard statistical techniques – Bayes theorem. The predictive distribution $\theta|\boldsymbol{X}$, posterior, has subjective interpretation (de Finetti's betting interpretation).
  - There is only one solution for each statistical problem. The remaining problem specific issue is to find the solution computationally and to select the right model + prior.

Bird's Eye of Statistical Inference

- Fiducial inference
  - Multiple distribution $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ are considered (no prior, but . . .).

Bird's Eye of Statistical Inference

- Fiducial inference
  - Multiple distribution $\mathcal{P} = \{P_\theta\}_{\theta\in\Theta}$ are considered (no prior, but . . .).
  - Goal is to find a distribution on the parameter space Θ that in summarizes the information we have obtained from the data.

Bird's Eye of Statistical Inference

- Fiducial inference
  - Multiple distribution $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ are considered (no prior, but . . .).
  - Goal is to find a distribution on the parameter space Θ that in summarizes the information we have obtained from the data.
  - Philosophical interpretation of fiducial probability is obscure.

Bird's Eye of Statistical Inference

- Fiducial inference
  - Multiple distribution $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ are considered (no prior, but . . .).
  - Goal is to find a distribution on the parameter space Θ that in summarizes the information we have obtained from the data.
  - Philosophical interpretation of fiducial probability is obscure.
  - We use fiducial distribution to propose statistical methods (e.g., confidence Intervals) and then evaluate the methods using repeated sampling performance.

Bird's Eye of Statistical Inference

- Fiducial inference
  - Multiple distribution $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ are considered (no prior, but . . .).
  - Goal is to find a distribution on the parameter space Θ that in summarizes the information we have obtained from the data.
  - Philosophical interpretation of fiducial probability is obscure.
  - We use fiducial distribution to propose statistical methods (e.g., confidence Intervals) and then evaluate the methods using repeated sampling performance.
  - The fiducial distribution is usually not a posterior with respect to any (data independent) prior (Grundy, 1956).

The aim of this talk

- We explain the definition of fiducial distribution as we generalize it demonstrating it on several examples.

The aim of this talk

- We explain the definition of fiducial distribution as we generalize it demonstrating it on several examples.
  - Applicable to both discrete and continuous distributions.

The aim of this talk

- We explain the definition of fiducial distribution as we generalize it demonstrating it on several examples.
  - Applicable to both discrete and continuous distributions.
  - We attempt to strip down all layers of additional structure.

The aim of this talk

- We explain the definition of fiducial distribution as we generalize it demonstrating it on several examples.
  - Applicable to both discrete and continuous distributions.
  - We attempt to strip down all layers of additional structure.
  - Our definition does not produce a "unique fiducial distribution". Regardless, the fiducial distribution is always proper.

The aim of this talk

- We explain the definition of fiducial distribution as we generalize it demonstrating it on several examples.
  - Applicable to both discrete and continuous distributions.
  - We attempt to strip down all layers of additional structure.
  - Our definition does not produce a "unique fiducial distribution". Regardless, the fiducial distribution is always proper.
- We proved some asymptotic theorems justifying this method of deriving inference procedures. Simulations usually show very good frequentist performance.

Comparison to MLE

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

Comparison to MLE

- 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 MLE

- Consider the data generating (structural) equation

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

  - U is a random variable/vector with known distribution
  - ξ is a fixed parameter.
  - The distribution of the data $\mathbf{X}$ is implied from U via the structural equation. I.e., one can generate $\mathbf{X}$ by generating U and plugging it into the structural equation.

Comparison to MLE

- Consider the data generating (structural) equation

$$\mathbf{x} = \mathbf{T}(U, \xi),$$

  - U is a random variable/vector with known distribution
  - ξ is a fixed parameter.
  - The distribution of the data $\mathbf{X}$ is implied from U via the structural equation.
- After observing $\mathbf{X} = \mathbf{x}$ deduce a distribution for ξ from that of U via the structural equation. I.e., generate ξ by generating $U^\star$, plugging it into the structural equation and solving for ξ.

Comparison to MLE

- Consider the data generating (structural) equation

$$\mathbf{x} = \mathbf{T}(U, \xi),$$

  - U is a random variable/vector with known distribution
  - ξ is a fixed parameter.
  - The distribution of the data $\mathbf{X}$ is implied from U via the structural equation.
- After observing $\mathbf{X} = \mathbf{x}$ deduce a distribution for ξ from that of U via the structural equation. I.e., generate ξ by generating $U^\star$, plugging it into the structural equation and solving for ξ.
- If the solution does not exist, discard this value of $U^\star$, i.e., condition the distribution of U on the fact that the solution exists.

Example – binomial

- Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p). Therefore

$$X_i = I_{[0,p)}(U_i), \quad i = 1, \ldots, n,$$

where U_i are i.i.d. uniform.

Example – binomial

- Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p). Therefore

$$X_i = I_{[0,p)}(U_i), \quad i = 1, \ldots, n,$$

  where U_i are i.i.d. uniform.

- Define the inverse image of T
  $Q(x_1, \ldots, x_n, u_1, \ldots, u_n) = \{p : x_i = I_{[0,p)}(u_i)\} = (m, M)$, where
  $m = \max_{x_i=1} u_i; M = \min_{x_i=0} u_i$

Example – binomial

- Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p). Therefore

$$X_i = I_{[0,p)}(U_i), \quad i = 1, \ldots, n,$$

where U_i are i.i.d. uniform.

- Define the inverse image of T
$Q(x_1, \ldots, x_n, u_1, \ldots, u_n) = \{p : x_i = I_{[0,p)}(u_i)\} = (m, M)$, where
$m = \max_{x_i=1} u_i$; $M = \min_{x_i=0} u_i$

- The fiducial distribution is

$$Q(x_1, \ldots, x_n, U_1^\star, \ldots, U_n^\star) \mid Q(x_1, \ldots, x_n, U_1^\star, \ldots, U_n^\star) \neq \emptyset$$
$$\stackrel{\mathcal{D}}{=} (U^\star_{(\sum x_i):n}, U^\star_{(1+\sum x_i):n})$$

Example – binomial

- Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p). Therefore

$$X_i = I_{[0,p)}(U_i), \quad i = 1, \ldots, n,$$

where U_i are i.i.d. uniform.

- Define the inverse image of T
$Q(x_1, \ldots, x_n, u_1, \ldots, u_n) = \{p : x_i = I_{[0,p)}(u_i)\} = (m, M)$, where
$m = \max_{x_i=1} u_i; M = \min_{x_i=0} u_i$

- The fiducial distribution is

$$Q(x_1, \ldots, x_n, U_1^\star, \ldots, U_n^\star) \mid Q(x_1, \ldots, x_n, U_1^\star, \ldots, U_n^\star) \neq \emptyset$$

$$\stackrel{\mathcal{D}}{=} (U^\star_{(\sum x_i):n}, U^\star_{(1+\sum x_i):n})$$

- We need to select a point inside the interval. We recommend selecting each edge with equal probability.

Single Normal

- Consider $X = \mu + Z$, where $Z \sim N(0, 1)$.

Single Normal

- Consider $X = \mu + Z$, where $Z \sim N(0, 1)$.
- Observe $X = 10$. Then we have $10 = \mu + Z$.

Single Normal

- Consider $X = \mu + Z$, where $Z \sim N(0, 1)$.
- Observe $X = 10$. Then we have $\mu = 10 - Z$.

Single Normal

- Consider $X = \mu + Z$, where $Z \sim N(0, 1)$.
- Observe $X = 10$. Then we have $\mu = 10 - Z$.
- Though the value of Z is unknown, we know the distribution of Z.

Single Normal

- Consider $X = \mu + Z$, where $Z \sim N(0, 1)$.
- Observe $X = 10$. Then we have $\mu = 10 - Z$.
- Though the value of Z is unknown, we know the distribution of Z.
- Fiducial argument:

$$P(\mu = 3 \pm dx) = P(10 - Z = 3 \pm dx) = P(Z = 7 \pm dx) \approx 1.83 \cdot 10^{-11} dx$$

Single Normal

- Consider $X = \mu + Z$, where $Z \sim N(0, 1)$.
- Observe $X = 10$. Then we have $\mu = 10 - Z$.
- Though the value of Z is unknown, we know the distribution of Z.
- Fiducial argument:

$$P(\mu = 3 \pm dx) = P(10 - Z = 3 \pm dx) = P(Z = 7 \pm dx) \approx 1.83 \cdot 10^{-11} dx$$

- This introduces a distribution on μ.

Single Normal

- Consider $X = \mu + Z$, where $Z \sim N(0,1)$.
- Observe $X = 10$. Then we have $\mu = 10 - Z$.
- Though the value of Z is unknown, we know the distribution of Z.
- Fiducial argument:

$$P(\mu = 3 \pm dx) = P(10 - Z = 3 \pm dx) = P(Z = 7 \pm dx) \approx 1.83 \cdot 10^{-11} dx$$

- This introduces a distribution on μ.
- We can simulate this distribution using $\mathcal{R}_\mu = 10 - Z^\star$, where $Z^\star \sim N(0,1)$ independent of Z.

Location Normal

- Consider $X_i = \mu + Z_i$ where Z_i are i.i.d. $N(0, 1)$.

Location Normal

- Consider $X_i = \mu + Z_i$ where Z_i are i.i.d. $N(0,1)$.
- Observe $(x_1, \ldots, x_n)$. We cannot simply follow the previous idea of setting $\mu = x_1 - Z_1^\star, \ldots, \mu = x_n - Z_n^\star$.

Location Normal

- Consider $X_i = \mu + Z_i$ where Z_i are i.i.d. $N(0,1)$.
- Observe $(x_1, \ldots, x_n)$. We cannot simply follow the previous idea of setting $\mu = x_1 - Z_1^\star, \ldots, \mu = x_n - Z_n^\star$. Each equation would lead to a different μ!

Location Normal

- Consider $X_i = \mu + Z_i$ where Z_i are i.i.d. $N(0,1)$.
- Observe $(x_1, \ldots, x_n)$. We cannot simply follow the previous idea of setting $\mu = x_1 - Z_1^\star, \ldots, \mu = x_n - Z_n^\star$. Each equation would lead to a different μ!
- Need to condition the distribution of $(Z_1^\star, \ldots, Z_n^\star)$ on the event that all the equations have the same μ.

Location Normal

- Consider $X_i = \mu + Z_i$ where Z_i are i.i.d. $N(0,1)$.
- Observe $(x_1, \ldots, x_n)$. We cannot simply follow the previous idea of setting $\mu = x_1 - Z_1^\star, \ldots, \mu = x_n - Z_n^\star$. Each equation would lead to a different μ!
- Need to condition the distribution of $(Z_1^\star, \ldots, Z_n^\star)$ on the event that all the equations have the same μ.
- The fiducial distribution can be defined as

$$x_1 - Z_1^\star | x_2 = x_1 - Z_1^\star + Z_2^\star, \ldots, x_n = x_1 - Z_1^\star + Z_n^\star.$$

Location Normal

- Consider $X_i = \mu + Z_i$ where Z_i are i.i.d. $N(0,1)$.
- Observe $(x_1, \ldots, x_n)$. We cannot simply follow the previous idea of setting $\mu = x_1 - Z_1^\star, \ldots, \mu = x_n - Z_n^\star$. Each equation would lead to a different μ!
- Need to condition the distribution of $(Z_1^\star, \ldots, Z_n^\star)$ on the event that all the equations have the same μ.
- The fiducial distribution can be defined as

$$x_1 - Z_1^\star | x_2 = x_1 - Z_1^\star + Z_2^\star, \ldots, x_n = x_1 - Z_1^\star + Z_n^\star.$$

- After simplification the fiducial distribution is $N(\bar{x}, 1/n)$.

Location Normal

- Consider $X_i = \mu + Z_i$ where Z_i are i.i.d. $N(0, 1)$.
- Observe $(x_1, \ldots, x_n)$. We cannot simply follow the previous idea of setting $\mu = x_1 - Z_1^\star, \ldots, \mu = x_n - Z_n^\star$. Each equation would lead to a different μ!
- Need to condition the distribution of $(Z_1^\star, \ldots, Z_n^\star)$ on the event that all the equations have the same μ.
- The fiducial distribution can be defined as

$$x_1 - Z_1^\star | x_2 = x_1 - Z_1^\star + Z_2^\star, \ldots, x_n = x_1 - Z_1^\star + Z_n^\star.$$

- After simplification the fiducial distribution is $N(\bar{x}, 1/n)$.
- We have non-uniqueness due to Borel paradox.

Remarks

- There are three challenges in the definition of generalized fiducial distribution.

Remarks

- There are three challenges in the definition of generalized fiducial distribution.
  - The choice of data generating equation. For i.i.d. data the $X_i = F^{-1}(U_i, \xi)$ is a good default choice.

Remarks

- There are three challenges in the definition of generalized fiducial distribution.
  - The choice of data generating equation. For i.i.d. data the $X_i = F^{-1}(U_i, \xi)$ is a good default choice.
  - The choice among multiple solutions:
    - Arises if the inverse image $Q(\mathbf{x}, U^\star)$ has more then one element but disappears asymptotically for parametric problems.

Remarks

- There are three challenges in the definition of generalized fiducial distribution.
  - The choice of data generating equation. For i.i.d. data the $X_i = F^{-1}(U_i, \xi)$ is a good default choice.
  - The choice among multiple solutions:
    - Arises if the inverse image $Q(\mathbf{x}, U^\star)$ has more then one element but disappears asymptotically for parametric problems.
  - The conditioning on the fact that solution exist:
    - Arises if $P\{Q(\mathbf{x}, U^\star) \neq \emptyset\} = 0$ – Borel paradox.
    - “Resolved by fat data”.

Fat data

- Borel paradox was caused by the fact that probability of observing our data could be 0.

Fat data

- Borel paradox was caused by the fact that probability of observing our data could be 0.
- Due to instrument limitations we never observe our data exactly.

Fat data

- Borel paradox was caused by the fact that probability of observing our data could be 0.
- Due to instrument limitations we never observe our data exactly.
  - My height is $1.85 < x_i < 1.86$.
  - Any number stored on a computer is known only to a machine precision.

Fat data

- Borel paradox was caused by the fact that probability of observing our data could be 0.
- Due to instrument limitations we never observe our data exactly.
  - My height is $1.85 < x_i < 1.86$.
  - Any number stored on a computer is known only to a machine precision.
- We derive generalized fiducial distribution directly for discretized data or take a limit as the discretization refines.

Fat data on a diet

- Assume that the data vector $\mathbf{x} \in \mathbb{R}^n$ has continuous distribution with the parameter $\xi \in \mathbb{R}^p$

Fat data on a diet

- Assume that the data vector $\mathbf{x} \in \mathbb{R}^n$ has continuous distribution with the parameter $\xi \in \mathbb{R}^p$
- Interpret fiducial recipe as the weak limit (as $\varepsilon \downarrow 0$) of

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

Fat data on a diet

- Assume that the data vector $\mathbf{x} \in \mathbb{R}^n$ has continuous distribution with the parameter $\xi \in \mathbb{R}^p$
- Interpret fiducial recipe as the weak limit (as $\varepsilon \downarrow 0$) of

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

  - The condition in (1) uses data fattened to a ball $\mathcal{B}_\varepsilon(\mathbf{x}) = \{\mathbf{y} : \|\mathbf{x} - \mathbf{y}\| < \varepsilon\}$

Fat data on a diet

- Assume that the data vector $\mathbf{x} \in \mathbb{R}^n$ has continuous distribution with the parameter $\xi \in \mathbb{R}^p$
- Interpret fiducial recipe as the weak limit (as $\varepsilon \downarrow 0$) of

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

  - The condition in (1) uses data fattened to a ball $\mathcal{B}_\varepsilon(\mathbf{x}) = \{\mathbf{y} : \|\mathbf{x} - \mathbf{y}\| < \varepsilon\}$
  - Similar to the idea of ABC; generating from prior replaced by $\min$.

Theoretical results

- If using $\| \ \|_\infty$ and smooth T the limiting conditional distribution (1) has density (Hannig, 2012)
$$r(\xi|\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x}|\xi)J(\mathbf{x},\xi)}{\int_\Xi f_{\mathbf{X}}(\mathbf{x}|\xi')J(\mathbf{x},\xi')\,d\xi'},$$

Theoretical results

- If using $\| \|_\infty$ and smooth T the limiting conditional distribution (1) has density (Hannig, 2012)

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

where $J(\mathbf{x},\xi) = \sum_{\mathbf{i}=(i_1,\ldots,i_p)} \left| \det \left(\frac{d}{d\xi}\mathbf{T}(\mathbf{u},\xi)\Big|_{\mathbf{u}=\mathbf{T}^{-1}(\mathbf{x},\xi)} \right)_{\mathbf{i}} \right|$

and $(A)_{\mathbf{i}}$ is the $p \times p$ matrix comprising of the $i_1, \ldots, i_p$th row of the $n \times p$ matrix A.

Comments

- Let $X_i = F^{-1}(\xi, U_i)$ be cont. with density $f(x|\xi)$.
  - Then $J(\mathbf{x}, \xi) = \sum_{\mathbf{i}=(i_1,\ldots,i_p)} \frac{\left|\det\left(\frac{d}{d\xi}\mathbf{F}(\mathbf{x_i},\xi)\right)\right|}{\prod_{\mathbf{i}} f(x_i,\xi)}$

Comments

- Let $X_i = F^{-1}(\xi, U_i)$ be cont. with density $f(x|\xi)$.
  - Then $J(\mathbf{x}, \xi) = \sum_{\mathbf{i}=(i_1,\ldots,i_p)} \frac{\left|\det\left(\frac{d}{d\xi}\mathbf{F}(\mathbf{x_i},\xi)\right)\right|}{\prod_{\mathbf{i}} f(x_i,\xi)}$
  - Often $\binom{n}{p}^{-1} J(\mathbf{x}, \xi) \to E_{\xi_0} \frac{\left|\det\left(\frac{d}{d\xi}\mathbf{F}(\mathbf{X_i},\xi)\right)\right|}{\prod_i f(x_i,\xi)}$ providing an empirical Bayes interpretation.

Comments

- Let $X_i = F^{-1}(\xi, U_i)$ be cont. with density $f(x|\xi)$.
  - Then $J(\mathbf{x}, \xi) = \sum_{\mathbf{i}=(i_1,\ldots,i_p)} \frac{\left|\det\left(\frac{d}{d\xi}\mathbf{F}(\mathbf{x_i},\xi)\right)\right|}{\prod_{\mathbf{i}} f(x_i,\xi)}$
  - Often $\binom{n}{p}^{-1} J(\mathbf{x}, \xi) \to E_{\xi_0} \frac{\left|\det\left(\frac{d}{d\xi}\mathbf{F}(\mathbf{X_i},\xi)\right)\right|}{\prod_i f(x_i,\xi)}$ providing an empirical Bayes interpretation.
- Confidence intervals based on generalized fiducial distribution are often correct asymptotically because of "Bernstein-von Mises" theorem for fiducial distributions Hannig (2009, 2012), Sonderegger & Hannig (2012).

Theoretical result for discretized data

- Assume structural equation $X_i = F^{-1}(U_i, \xi)$
  - ξ is p dimensional and U_i are i.i.d. $U(0,1)$.
  - $F(x, \xi)$ is continuously differentiable in ξ for all x
  - $(F(x_1, \xi), \ldots, F(x_p, \xi)) = (u_1, \ldots, u_p)$, taken as a function of ξ is one to one for each $\mathbf{x}$.

Theoretical result for discretized data

- Assume structural equation $X_i = F^{-1}(U_i, \xi)$
  - ξ is p dimensional and U_i are i.i.d. $U(0,1)$.
  - $F(x, \xi)$ is continuously differentiable in ξ for all x
  - $(F(x_1, \xi), \ldots, F(x_p, \xi)) = (u_1, \ldots, u_p)$, taken as a function of ξ is one to one for each $\mathbf{x}$.
- Data were discretized to a fixed partition $(-\infty, a_1], (a_1, a_2], \ldots, (a_k, \infty)$.
  - $P(X \in (a_j, a_{j+1}]) > 0$ for all j.
  - For all $\mathbf{j} \subset \{1, \ldots, k\}$, the Jacobian $\det \left(\frac{\mathbf{d} F(\mathbf{a_j}, \xi_0)}{\mathbf{d}\xi} \right) \neq 0$.

Theoretical result for discretized data

- Assume structural equation $X_i = F^{-1}(U_i, \xi)$
  - ξ is p dimensional and U_i are i.i.d. $U(0,1)$.
  - $F(x,\xi)$ is continuously differentiable in ξ for all x
  - $(F(x_1,\xi),\dots,F(x_p,\xi)) = (u_1,\dots,u_p)$, taken as a function of ξ is one to one for each $\mathbf{x}$.
- Data were discretized to a fixed partition $(-\infty, a_1], (a_1, a_2], \dots, (a_k, \infty)$.
  - $P(X \in (a_j, a_{j+1}]) > 0$ for all j.
  - For all $\mathbf{j} \subset \{1,\dots,k\}$, the Jacobian $\det\left(\frac{\mathbf{d}F(\mathbf{a_j},\xi_0)}{\mathbf{d}\xi}\right) \neq 0$.

Theorem (Hannig (2012)). *Confidence sets based on the generalized fiducial distribution will have asymptotically correct coverage as number of data points goes to infinity and resolution remains fixed.*

Model Selection

- Consider several models $(\mathcal{M}_i)_{i \in I}$, each with data generating equation $\boldsymbol{X} = T_i(\mathbf{U}, \xi), i \in I$.

Model Selection

- Consider several models $(\mathcal{M}_i)_{i \in I}$, each with data generating equation $\boldsymbol{X} = T_i(\mathbf{U}, \xi), i \in I$.
- If all models have the same dimension (number of parameters), the marginal distribution of ith model is

$$r(\mathcal{M}_i) \propto \int_{\Xi} f_i(\mathbf{x}|\xi) J_i(\mathbf{x}, \xi) \, d\xi,$$

where f_i and J_i were computed using the ith model.

Model Selection

- Consider several models $(\mathcal{M}_i)_{i \in I}$, each with data generating equation $\boldsymbol{X} = T_i(\mathbf{U}, \xi), i \in I$.
- If all models have the same dimension (number of parameters), the marginal distribution of ith model is

$$r(\mathcal{M}_i) \propto \int_\Xi f_i(\mathbf{x}|\xi) J_i(\mathbf{x}, \xi)\, d\xi,$$

where f_i and J_i were computed using the ith model.

- When the number of parameters is different, penalty is needed.

$$r(\mathcal{M}_i) \propto e^{-q(i)} \int_\Xi f_i(\mathbf{x}|\xi) J_i(\mathbf{x}, \xi)\, d\xi;$$

we choose the Minimum Description Length penalty for $q(i)$.

Model Selection

- Consider several models $(\mathcal{M}_i)_{i\in I}$, each with data generating equation $\boldsymbol{X} = T_i(\mathbf{U}, \xi), i \in I$.
- If all models have the same dimension (number of parameters), the marginal distribution of ith model is

$$r(\mathcal{M}_i) \propto \int_{\Xi} f_i(\mathbf{x}|\xi) J_i(\mathbf{x}, \xi)\, d\xi,$$

  where f_i and J_i were computed using the ith model.
- When the number of parameters is different, penalty is needed.

$$r(\mathcal{M}_i) \propto e^{-q(i)} \int_{\Xi} f_i(\mathbf{x}|\xi) J_i(\mathbf{x}, \xi)\, d\xi;$$

  we choose the Minimum Description Length penalty for $q(i)$.
- Wandler & Hannig (2011, 2012) shows consistency for various multivariate normal model.

Key comparison

- National laboratories of various countries e.g, NIST, NPL, carry out interlaboratory trials to evaluate the relative measurement capabilities of each other and also establish standard reference values.

Key comparison

- National laboratories of various countries e.g, NIST, NPL, carry out interlaboratory trials to evaluate the relative measurement capabilities of each other and also establish standard reference values.
- Fo simplicity assume, each laboratory reports a confidence interval based on a T distribution, measuring the same object.

Key comparison

- National laboratories of various countries e.g, NIST, NPL, carry out interlaboratory trials to evaluate the relative measurement capabilities of each other and also establish standard reference values.
- Fo simplicity assume, each laboratory reports a confidence interval based on a T distribution, measuring the same object.
- Goal is to combine the intervals in a way that down ways potential outliers. Outright dropping of odd results is politically not feasible.

Key comparison - our solution

- All possible subsets of laboratories are considered as measuring the correct values with the rest considered as outliers.

Key comparison - our solution

- All possible subsets of laboratories are considered as measuring the correct values with the rest considered as outliers.
- For each model the fiducial distributions measuring the true value are combined using Hannig & Xie (2012) and the fiducial model probability is computed.

$$r(\mu) \propto \sum_{j \in I} C(\mathcal{M}_j) \sum_{i \in \mathcal{M}_j} \left\{ \frac{1}{n_i} + \frac{(\mu - \bar{x}_i)^2}{(n_i - 1)s_i^2} \right\}^{-1/2} \prod_{i \in \mathcal{M}_j} \left\{ 1 + \frac{n_i(\mu - \bar{x}_i)^2}{(n_i - 1)s_i^2} \right\}^{-(n_i-1)/2} \times e^{-(k-|\mathcal{M}_i||)\log(SSE)/2}$$

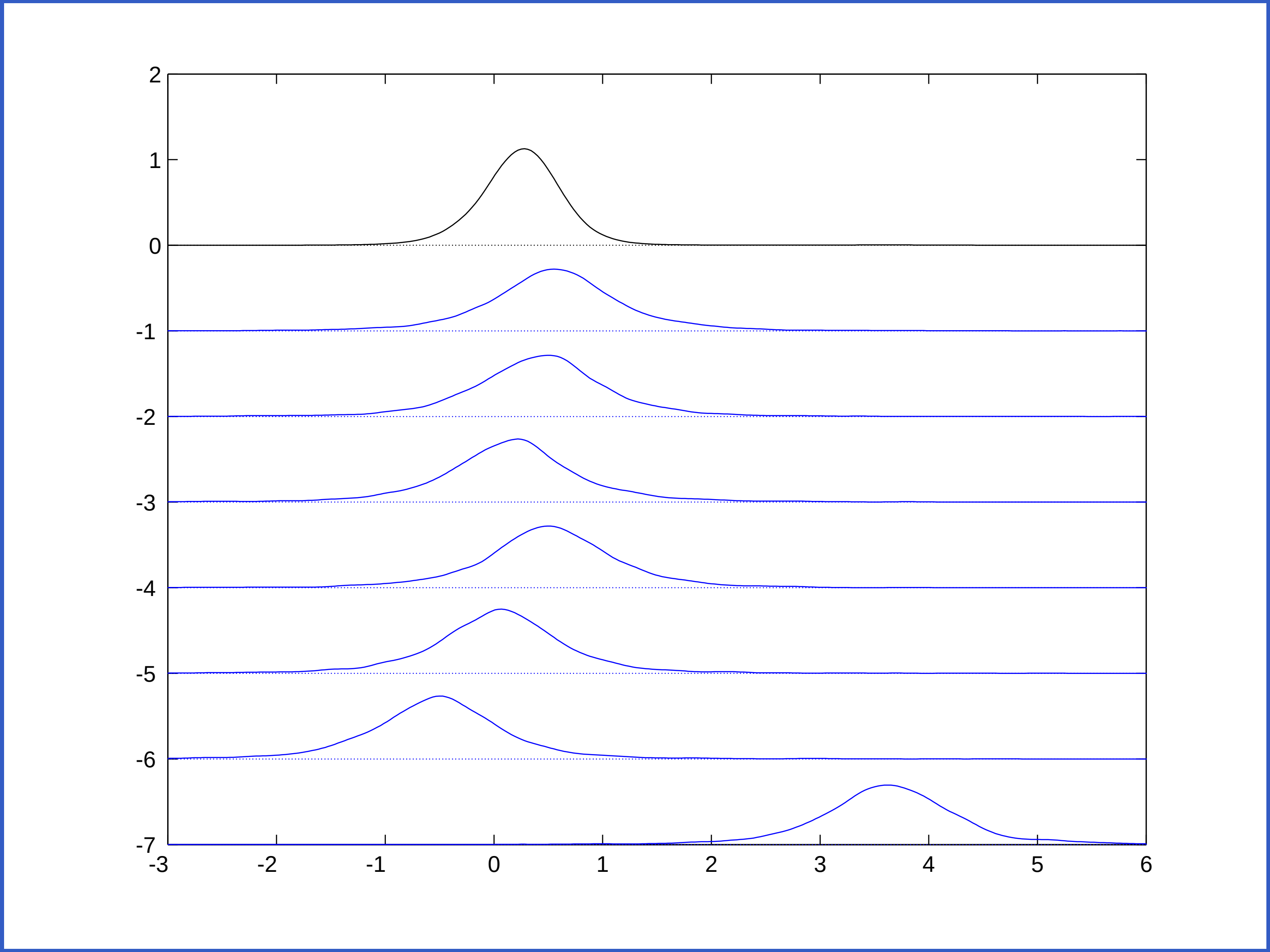
Key comparison - our solution

- All possible subsets of laboratories are considered as measuring the correct values with the rest considered as outliers.
- For each model the fiducial distributions measuring the true value are combined using Hannig & Xie (2012) and the fiducial model probability is computed.

$$r(\mu) \propto \sum_{j\in I} C(\mathcal{M}_j) \sum_{i\in\mathcal{M}_j} \left\{ \frac{1}{n_i} + \frac{(\mu-\bar{x}_i)^2}{(n_i-1)s_i^2} \right\}^{-1/2} \prod_{i\in\mathcal{M}_j} \left\{ 1 + \frac{n_i(\mu-\bar{x}_i)^2}{(n_i-1)s_i^2} \right\}^{-(n_i-1)/2}$$
$$\times e^{-(k-|\mathcal{M}_i||)\log(SSE)/2}$$

- Preliminary simulation using importance sampling shows somewhat conservative performance.

Key comparison - example



Example - Linear Mixed Model

- Modern linear mixed models can be traced back to Fisher (1935) and Bartlet (1937).

Example - Linear Mixed Model

- Modern linear mixed models can be traced back to Fisher (1935) and Bartlet (1937).
- Point estimation has a unified approach that works well and is widely used in practice (REML) Patterson & Thompson (1971).

Example - Linear Mixed Model

- Modern linear mixed models can be traced back to Fisher (1935) and Bartlet (1937).
- Point estimation has a unified approach that works well and is widely used in practice (REML) Patterson & Thompson (1971).
- There seems to be no (non-Bayesian) unified approach producing good quality confidence sets. Most procedures in the literature are designed to solve special cases Burdick, Graybill, Wang or use insufficient statistics Khuri, Mathews and Sinha.

Example - Linear Mixed Model

- Modern linear mixed models can be traced back to Fisher (1935) and Bartlet (1937).
- Point estimation has a unified approach that works well and is widely used in practice (REML) Patterson & Thompson (1971).
- There seems to be no (non-Bayesian) unified approach producing good quality confidence sets. Most procedures in the literature are designed to solve special cases Burdick, Graybill, Wang or use insufficient statistics Khuri, Mathews and Sinha.
- We will propose a procedure that produces confidence sets for large class of linear mixed models. Additionally it allows for discretized data.

Linear Mixed Model

- Consider a structural equation

$$\mathbf{Y} = \mathbf{X}\beta + \sum_{i=1}^{k} \sigma_i \sum_{j=1}^{l_k} V_{i,j} Z_{i,j}$$

Linear Mixed Model

- Consider a structural equation

$$\mathbf{Y} = \mathbf{X}\beta + \sum_{i=1}^{k} \sigma_i \sum_{j=1}^{l_k} V_{i,j} Z_{i,j}$$

  - $\mathbf{Y}$ observations, $\mathbf{X}$ design matrix, β fixed effect parameters
  - k number of random effects, l_k number of levels per effect,
  - $V_{i,j}$ var component design vectors, σ_i^2 variance of the ith effect
  - $Z_{i,j}$ i.i.d. $N(0,1)$

Linear Mixed Model

- Consider a structural equation

$$\mathbf{Y} = \mathbf{X}\beta + \sum_{i=1}^{k} \sigma_i \sum_{j=1}^{l_k} V_{i,j} Z_{i,j}$$

  - $\mathbf{Y}$ observations, $\mathbf{X}$ design matrix, β fixed effect parameters
  - k number of random effects, l_k number of levels per effect,
  - $V_{i,j}$ var component design vectors, σ_i^2 variance of the ith effect
  - $Z_{i,j}$ i.i.d. $N(0, 1)$
- Contains a wide variety of linear mixed models.

Linear Mixed Model

$$\mathbf{Y} = \mathbf{X}\beta + \sum_{i=1}^{k} \sigma_i \sum_{j=1}^{l_k} V_{i,j} Z_{i,j}$$

- Linear regression
  - $k = 1$, $l_1 = n$, $V_{1,\cdot} = (V_{1,1}, \ldots, V_{1,n}) = I$
  - m regression coefficients, σ_1^2 error variance

Linear Mixed Model

$$\mathbf{Y} = \mathbf{X}\beta + \sum_{i=1}^{k} \sigma_i \sum_{j=1}^{l_k} V_{i,j} Z_{i,j}$$

- Linear regression
  - $k = 1$, $l_1 = n$, $V_{1,\cdot} = (V_{1,1}, \ldots, V_{1,n}) = I$
  - m regression coefficients, σ_1^2 error variance
- One way random effects model
  - $X = 1$, $k = 2$, l_1 number of levels for random effect, $l_2 = n$, $V_{1,i}$ indicates which observations are in group i, $V_{2,\cdot} = I$
  - m overall mean, σ_1^2 random effect variance, σ_2^2 error variance

Linear Mixed Model

- Assume we observe $\mathbf{a} \leq \mathbf{Y} \leq \mathbf{b}$

Linear Mixed Model

- Assume we observe $\mathbf{a} \leq \mathbf{Y} \leq \mathbf{b}$
- We can generate $Z^{\star}_{i,j}$ as i.i.d. N(0,1) and solve for β, σ – linear programming problem

Linear Mixed Model

- Assume we observe $\mathbf{a} \leq \mathbf{Y} \leq \mathbf{b}$
- We can generate $Z^{\star}_{i,j}$ as i.i.d. N(0,1) and solve for β, σ – linear programming problem
- The conditioning implies, we need to exclude all $Z^{\star}_{i,j}$ for which there is no solution.

Linear Mixed Model

- Assume we observe $\mathbf{a} \le \mathbf{Y} \le \mathbf{b}$
- We can generate $Z^{\star}_{i,j}$ as i.i.d. N(0,1) and solve for β, σ – linear programming problem
- The conditioning implies, we need to exclude all $Z^{\star}_{i,j}$ for which there is no solution.
- Need an efficient algorithm for generating such $\boldsymbol{Z}$.

Linear Mixed Model

- Assume we observe $\mathbf{a} \leq \mathbf{Y} \leq \mathbf{b}$
- We can generate $Z^{\star}_{i,j}$ as i.i.d. N(0,1) and solve for β, σ – linear programming problem
- The conditioning implies, we need to exclude all $Z^{\star}_{i,j}$ for which there is no solution.
- Need an efficient algorithm for generating such $\boldsymbol{Z}$.
- Possibilities include
  - Gibbs sampler – does not mix well if there is too much precision.
  - Simulated tampering – works but slow
  - We proposed a particular implementation of Sequential Monte Carlo algorithm – works well if the number of parameters is reasonable (< 10).

Simulation study

- **One-way random effects:** $Y_{ijk} = \mu + \alpha_i + \epsilon_{ij}$
  (μ fixed, α and ϵ are independent and $\sim$ Normal)

Simulation study

- **One-way random effects:** $Y_{ijk} = \mu + \alpha_i + \epsilon_{ij}$
  (μ fixed, α and ϵ are independent and $\sim$ Normal)
- **Two-fold nested:** $Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk}$
  (μ fixed, α, β and ϵ are independent and $\sim$ Normal)

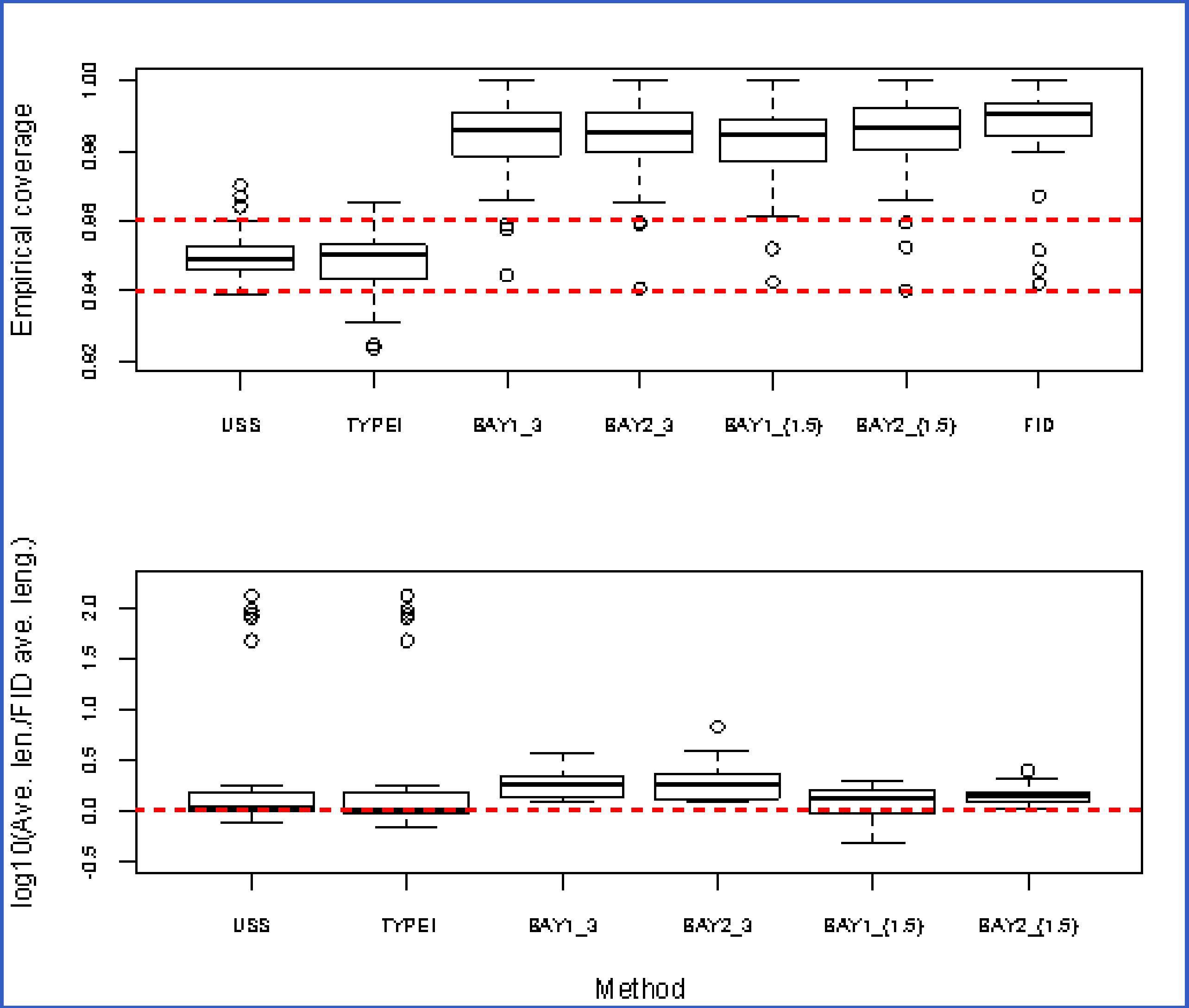
Simulation study

- **One-way random effects:** $Y_{ijk} = \mu + \alpha_i + \epsilon_{ij}$
  (μ fixed, α and ϵ are independent and $\sim$ Normal)
- **Two-fold nested:** $Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk}$
  (μ fixed, α, β and ϵ are independent and $\sim$ Normal)
- **Two-factor crossed design with interaction:**
  $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
  (μ fixed, α, β, $(\alpha\beta)$, and ϵ are independent and $\sim$ Normal)

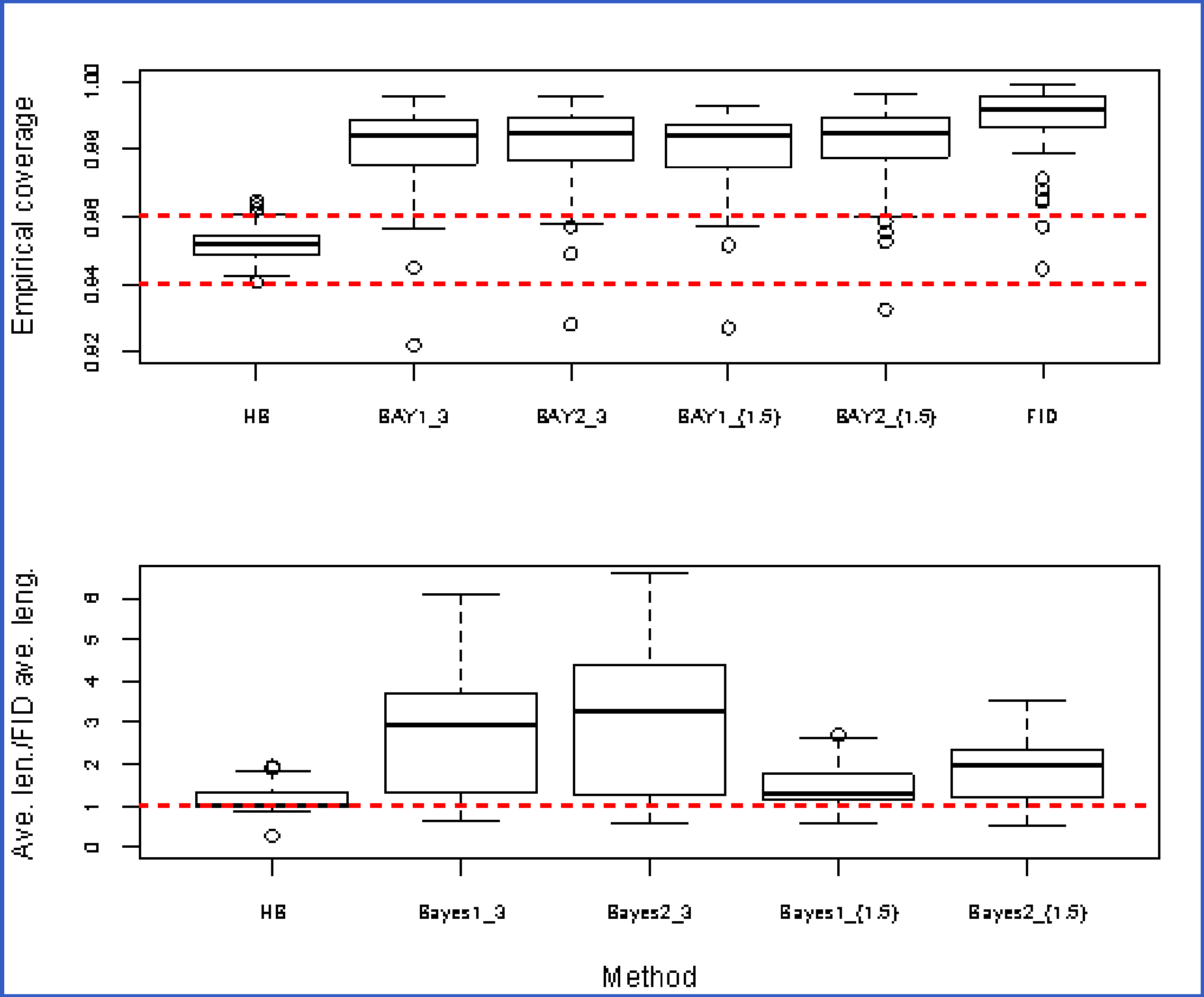
Simulation study

- **One-way random effects:** $Y_{ijk} = \mu + \alpha_i + \epsilon_{ij}$
  (μ fixed, α and ϵ are independent and $\sim$ Normal)
- **Two-fold nested:** $Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk}$
  (μ fixed, α, β and ϵ are independent and $\sim$ Normal)
- **Two-factor crossed design with interaction:**
  $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
  (μ fixed, α, β, $(\alpha\beta)$, and ϵ are independent and $\sim$ Normal)
- We considered a number of models with various levels of imbalance and values of parameters.

95% CI for random effects (nested)



95% CI for random effects (crossed)



Some generalized fiducial projects

We applied generalized fiducial inference to:

Some generalized fiducial projects

We applied generalized fiducial inference to:

- Unbalanced random effect model with amazing results
  + Linear Mixed Model

Some generalized fiducial projects

We applied generalized fiducial inference to:

- Unbalanced random effect model with amazing results
  + Linear Mixed Model
- Logistic regression (LD50) + General Linear Models.

Some generalized fiducial projects

We applied generalized fiducial inference to:

- Unbalanced random effect model with amazing results + Linear Mixed Model
- Logistic regression (LD50) + General Linear Models.
- Confidence sets for wavelet regression using fiducial idea. This is related to model selection.

Some generalized fiducial projects

We applied generalized fiducial inference to:

- Unbalanced random effect model with amazing results + Linear Mixed Model
- Logistic regression (LD50) + General Linear Models.
- Confidence sets for wavelet regression using fiducial idea. This is related to model selection.
- Extreme value data (Generalized Pareto) & Maximum mean (QT intervals) and model comparison.

Some generalized fiducial projects

We applied generalized fiducial inference to:

- Unbalanced random effect model with amazing results + Linear Mixed Model
- Logistic regression (LD50) + General Linear Models.
- Confidence sets for wavelet regression using fiducial idea. This is related to model selection.
- Extreme value data (Generalized Pareto) & Maximum mean (QT intervals) and model comparison.
- Ultra-highdimensional Regression Model (How to properly introduce a penalty?)

Concluding remarks

- Generalized fiducial distributions lead often to attractive solution with asymptotically correct frequentist coverage.

Concluding remarks

- Generalized fiducial distributions lead often to attractive solution with asymptotically correct frequentist coverage.
- Many simulation studies show that generalized fiducial solutions have very good small sample properties.

Concluding remarks

- Generalized fiducial distributions lead often to attractive solution with asymptotically correct frequentist coverage.
- Many simulation studies show that generalized fiducial solutions have very good small sample properties.
- Current popularity of generalized inference in some applied circles suggests that if computers were available 70 years ago, fiducial inference might not have been rejected.

Quotes

- Zabell (1992) "*Fiducial inference stands as R. A. Fisher's one great failure.*"
- Efron (1998) "*Maybe Fisher's biggest blunder will become a big hit in the 21st century!*"

Quotes

- Zabell (1992) "*Fiducial inference stands as R. A. Fisher's one great failure.*"
- Efron (1998) "*Maybe Fisher's biggest blunder will become a big hit in the 21st century!*"