On Generalized Fiducial Inference

Jan Hannig

jan.hannig@colostate.edu

University of North Carolina at Chapel Hill

Parts of this talk are based on joint work with:

Hari Iyer, Thomas C.M. Lee,

Paul Patterson, Lidong E, Damian Wandler and Derek Sonderegger

Colorado State University

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

Fiducial: A Brief History

- goal: to construct distributions for model parameters

Fiducial: A Brief History

- goal: to construct distributions for model parameters
- introduced by Fisher (1930, 1935)

Fiducial: A Brief History

- goal: to construct distributions for model parameters
- introduced by Fisher (1930, 1935)
- attempt to overcome what he saw as an issue of the Bayesian approach to inference

Fiducial: A Brief History

- goal: to construct distributions for model parameters
- introduced by Fisher (1930, 1935)
- attempt to overcome what he saw as an issue of the Bayesian approach to inference – use of a prior distribution when no prior information was available

Fiducial: A Brief History

- goal: to construct distributions for model parameters
- introduced by Fisher (1930, 1935)
- attempt to overcome what he saw as an issue of the Bayesian approach to inference – use of a prior distribution when no prior information was available
- related work: Fraser (1960), Dawid and Stone (1982), Dempster (1968, 2008).

Fiducial: A Brief History

- goal: to construct distributions for model parameters
- introduced by Fisher (1930, 1935)
- attempt to overcome what he saw as an issue of the Bayesian approach to inference – use of a prior distribution when no prior information was available
- related work: Fraser (1960), Dawid and Stone (1982), Dempster (1968, 2008).
- it is fair to say that fiducial inference failed to occupy an important place in mainstream statistics

Recent Developments

- Weerahandi (1989, 1993) proposed new concepts of generalized confidence intervals

Recent Developments

- Weerahandi (1989, 1993) proposed new concepts of generalized confidence intervals
- Hannig, Iyer & Patterson (2006) noted that every published generalized confidence interval was obtainable using the fiducial arguments

Recent Developments

- Weerahandi (1989, 1993) proposed new concepts of generalized confidence intervals
- Hannig, Iyer & Patterson (2006) noted that every published generalized confidence interval was obtainable using the fiducial arguments
- and they proved the asymptotic frequentist correctness of such intervals

Recent Developments

- Weerahandi (1989, 1993) proposed new concepts of generalized confidence intervals
- Hannig, Iyer & Patterson (2006) noted that every published generalized confidence interval was obtainable using the fiducial arguments
- and they proved the asymptotic frequentist correctness of such intervals
- Hannig (2008) have developed/modified these ideas further — termed the resulting work generalized fiducial inference

What was on Fisher's mind?

What was on Fisher's mind?

- a "Switching Principle"

What was on Fisher's mind?

- a "Switching Principle"
- for the celebrated Maximum Likelihood method

What was on Fisher's mind?

- a "Switching Principle"
- for the celebrated Maximum Likelihood method
  - *density* is $f(\boldsymbol{X}, \theta)$, where $\boldsymbol{X}$ is variable and θ is fixed

What was on Fisher's mind?

- a "Switching Principle"
- for the celebrated Maximum Likelihood method
  - *density* is $f(\boldsymbol{X}, \theta)$, where $\boldsymbol{X}$ is variable and θ is fixed
  - *likelihood* is $f(\boldsymbol{X}, \theta)$, where $\boldsymbol{X}$ is fixed and θ is variable

What was on Fisher's mind?

- a "Switching Principle"
- for the celebrated Maximum Likelihood method
  - *density* is $f(\boldsymbol{X}, \theta)$, where $\boldsymbol{X}$ is variable and θ is fixed
  - *likelihood* is $f(\boldsymbol{X}, \theta)$, where $\boldsymbol{X}$ is fixed and θ is variable
- as we will see, generalized fiducial inference is also based on this idea

What was on Fisher's mind?

- a "Switching Principle"
- for the celebrated Maximum Likelihood method
  - *density* is $f(\boldsymbol{X}, \theta)$, where $\boldsymbol{X}$ is variable and θ is fixed
  - *likelihood* is $f(\boldsymbol{X}, \theta)$, where $\boldsymbol{X}$ is fixed and θ is variable
- as we will see, generalized fiducial inference is also based on this idea
- the switching of the roles of $\boldsymbol{X}$ and θ

Simplistic Example 1

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

Simplistic Example 1

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

Simplistic Example 1

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

Simplistic Example 1

- 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.

Simplistic Example 1

- 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:

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

Simplistic Example 1

- 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.

Simplistic Example 1

- 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.

Simplistic Example 2

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

Simplistic Example 2

- 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 $\mu = x_1 - Z_1^\star, \ldots, \mu = x_n - Z_n^\star$.

Simplistic Example 2

- 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!

Simplistic Example 2

- 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.

Simplistic Example 2

- 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

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

Simplistic Example 2

- 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 $N(\bar{x}, 1/n)$.

Simplistic Example 2

- 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.

Example 3—“Fat data”

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

Example 3—“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.

Example 3—"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.
  - We only know that they are in some interval, i.e., we do not observe $x_i = \pi$ but rather $3.141 < x_i < 3.142$.

Example 3—“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.
  - We only know that they are in some interval, i.e., we do not observe $x_i = \pi$ but rather $3.141 < x_i < 3.142$.
- Let $\mathbf{X} = \mu + \sigma\mathbf{Z}$.

Example 3—“Fat data”

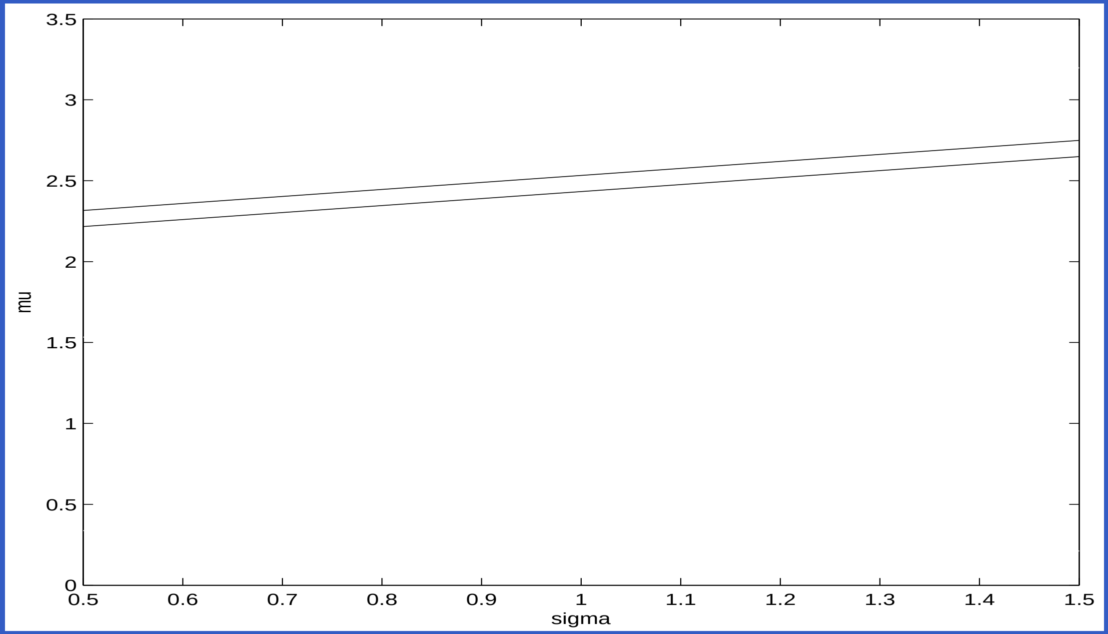
- Borel paradox was caused by the fact that probability of observing our data could be 0.
- Due to instrument limitations we never observe our data exactly.
  - We only know that they are in some interval, i.e., we do not observe $x_i = \pi$ but rather $3.141 < x_i < 3.142$.
- Let $\mathbf{X} = \mu + \sigma \mathbf{Z}$.
  - If we observe $a_i < X_i < b_i$ we need to generate $\mathbf{Z}^\star$ keeping only those values that agree with $a_i < \mu + \sigma \mathbf{Z}_i^\star < b_i$ for all i.

Example 3—“Fat data”

- Say we observe $2.0 < X_1 < 2.1,\ 0.6 < X_2 < 0.7,\ 0.4 < X_3 < 0.5$.

Example 3—“Fat data”

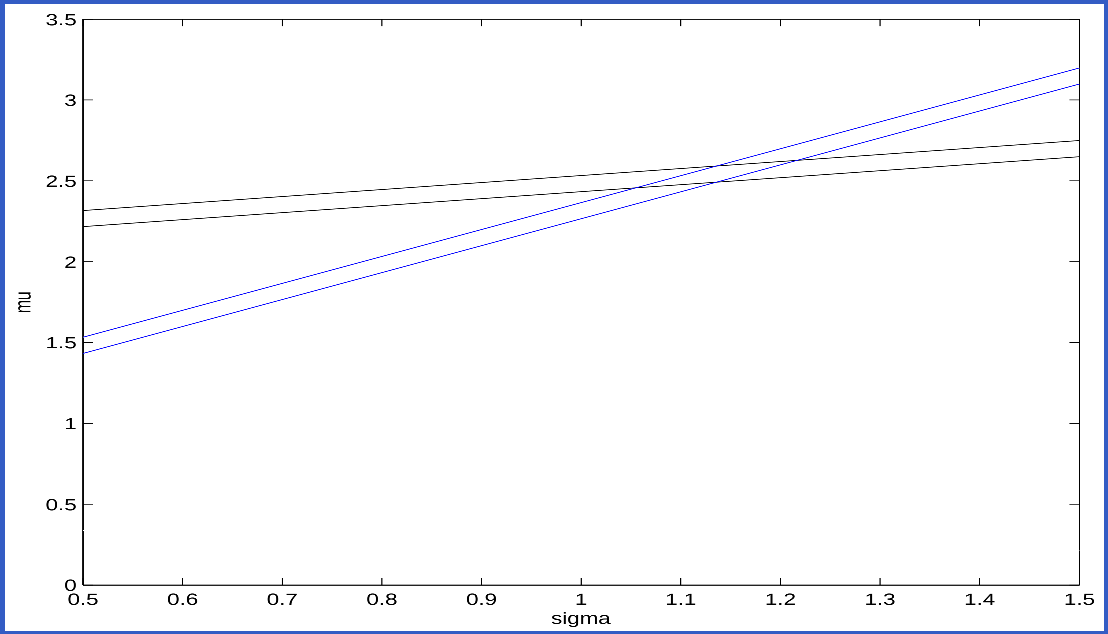
- Say we observe $2.0 < X_1 < 2.1, \ 0.6 < X_2 < 0.7, \ 0.4 < X_3 < 0.5$.



$$Z_1^\star = -0.4326$$

Example 3—“Fat data”

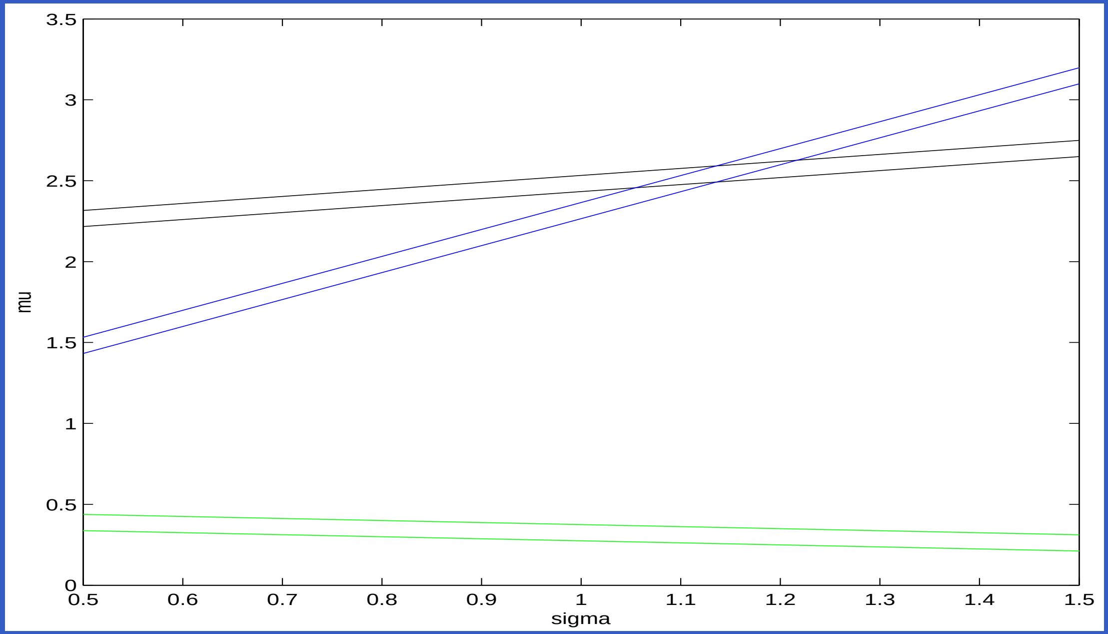
- Say we observe $2.0 < X_1 < 2.1,\ 0.6 < X_2 < 0.7,\ 0.4 < X_3 < 0.5$.



$$Z_1^\star = -0.4326, Z_2^\star = -1.6656$$

Example 3—"Fat data"

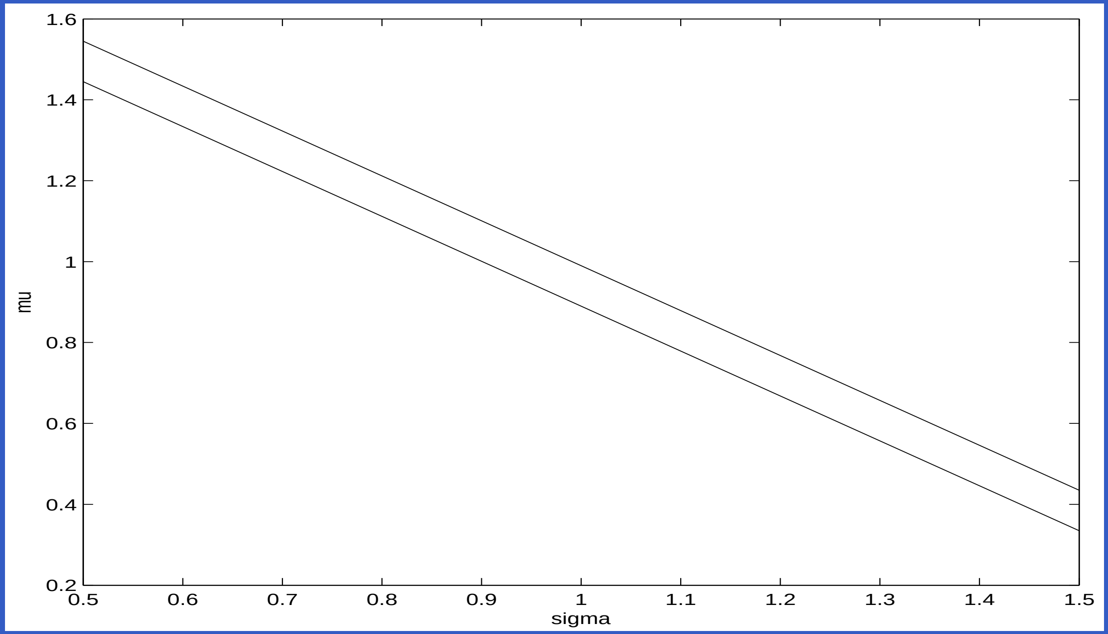
- Say we observe $2.0 < X_1 < 2.1,\ 0.6 < X_2 < 0.7,\ 0.4 < X_3 < 0.5$.



$$Z_1^\star = -0.4326, Z_2^\star = -1.6656, Z_3^\star = 0.1253$$

Example 3—"Fat data"

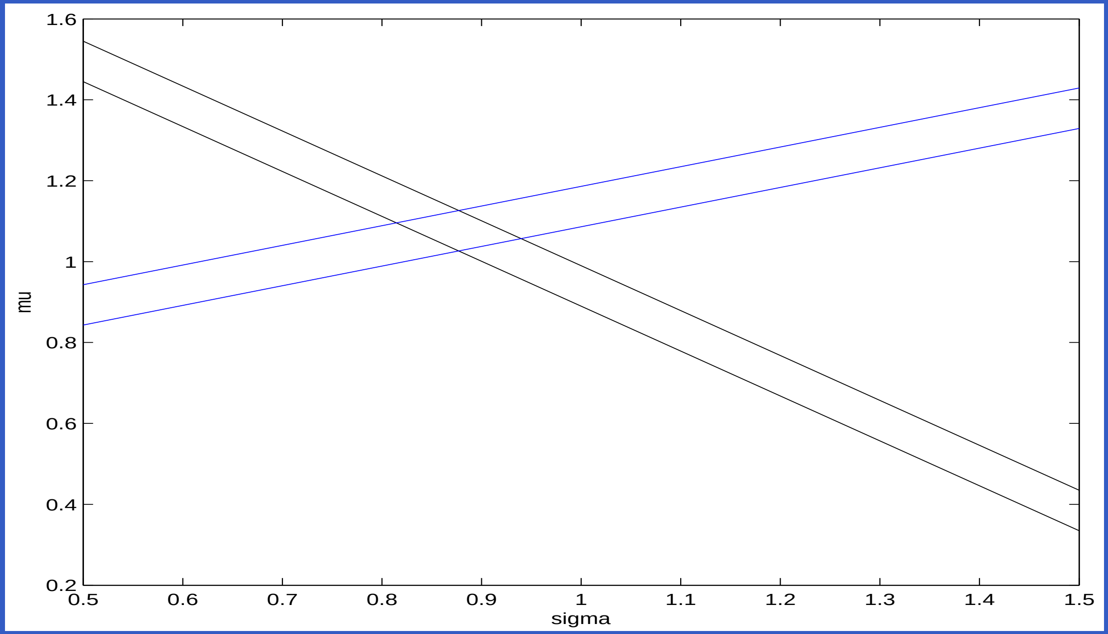
- Say we observe $2.0 < X_1 < 2.1,\ 0.6 < X_2 < 0.7,\ 0.4 < X_3 < 0.5$.



$$Z_1^\star = 1.102$$

Example 3—“Fat data”

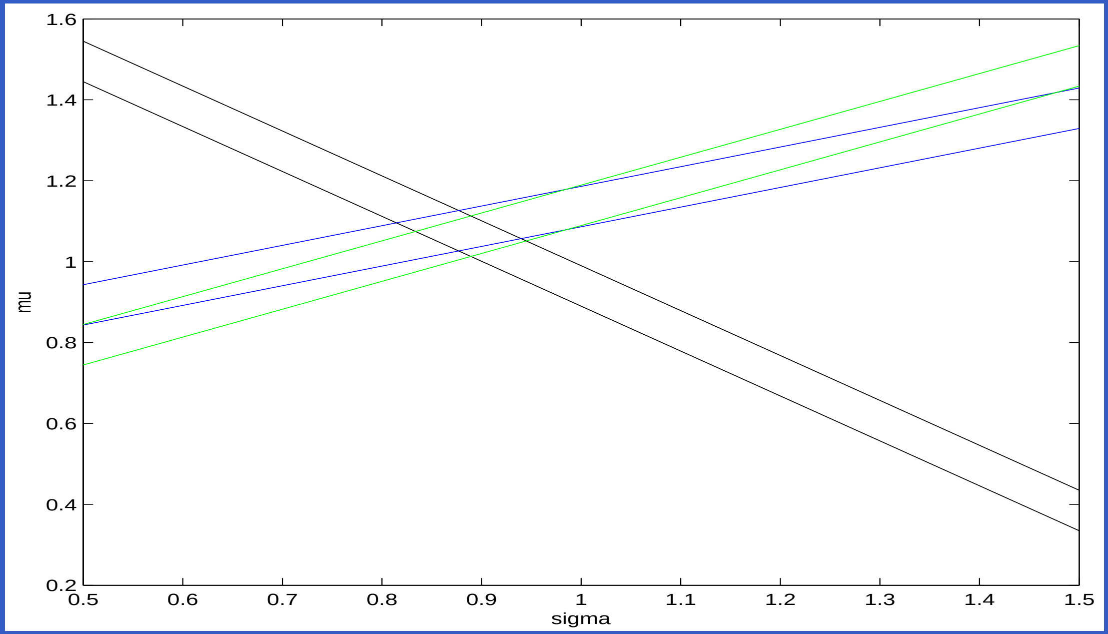
- Say we observe $2.0 < X_1 < 2.1,\ 0.6 < X_2 < 0.7,\ 0.4 < X_3 < 0.5$.



$$Z_1^\star = 1.1102, Z_2^\star = -0.4861$$

Example 3—“Fat data”

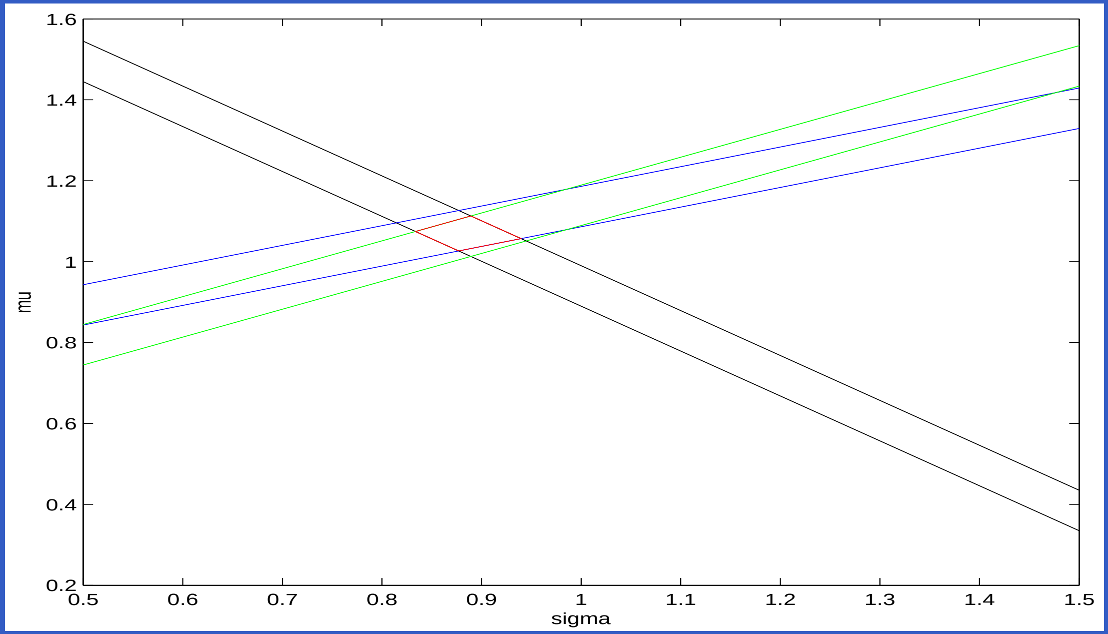
- Say we observe $2.0 < X_1 < 2.1,\ 0.6 < X_2 < 0.7,\ 0.4 < X_3 < 0.5$.



$$Z_1^\star = 1.1102, Z_2^\star = -0.4861, Z_3^\star = -0.6892$$

Example 3—“Fat data”

- Say we observe $2.0 < X_1 < 2.1,\ 0.6 < X_2 < 0.7,\ 0.4 < X_3 < 0.5$.



$$Z_1^\star = 1.1102, Z_2^\star = -0.4861, Z_3^\star = -0.6892$$

- Denote the intersection by Q.

Example 3—"Fat data"

- Set $Q((\mathbf{a},\mathbf{b}),u) = \{(\mu,\sigma) : a_i < \mu + \sigma \mathbf{Z}_i^\star < b_i\}$.

Example 3—"Fat data"

- Set $Q((\mathbf{a},\mathbf{b}),u) = \{(\mu,\sigma) : a_i < \mu + \sigma \mathbf{Z}_i^\star < b_i\}$.
- The fiducial distribution could be defined as $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \neq \emptyset\}$

Example 3—"Fat data"

- Set $Q((\mathbf{a},\mathbf{b}),u) = \{(\mu,\sigma) : a_i < \mu + \sigma \mathbf{Z}_i^\star < b_i\}$.
- The fiducial distribution could be defined as $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \neq \emptyset\}$
- $P(Q \neq \emptyset) \geq P(\mathbf{X} \in (\mathbf{a},\mathbf{b})) > 0$, so there is no Borel paradox in the definition of fiducial distribution (1).

Example 3—“Fat data”

- Set $Q((\mathbf{a},\mathbf{b}),u) = \{(\mu,\sigma) : a_i < \mu + \sigma \mathbf{Z}_i^\star < b_i\}$.
- The fiducial distribution could be defined as $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \neq \emptyset\}$
- $P(Q \neq \emptyset) \geq P(\mathbf{X} \in (\mathbf{a},\mathbf{b})) > 0$, so there is no Borel paradox in the definition of fiducial distribution (1).
- Q typically contains more than one element. We can either use Dempster-Shafer calculus to interpret its meaning, or additionally choose (randomly) an element from Q.

Example 3—"Fat Data"

- Let $\mathbf{X} = \mu + \sigma\mathbf{Z}$. We observe (2.0,2.1), (0.6,0.7), (0.4, 0.5), (1.4,1.5), (0.7,0.8), (0.8,0.9), (1.2,1.3), (1.2,1.3), (1.1,1.2), (1.5,1.6), (1.4,1.5), (0.4,0.5), (1.2,1.3), (0.7,0.8), (0.5,0.6).

Example 3—"Fat Data"

- Let $\mathbf{X} = \mu + \sigma\mathbf{Z}$. We observe (2.0,2.1), (0.6,0.7), (0.4, 0.5), (1.4,1.5), (0.7,0.8), (0.8,0.9), (1.2,1.3), (1.2,1.3), (1.1,1.2), (1.5,1.6), (1.4,1.5), (0.4,0.5), (1.2,1.3), (0.7,0.8), (0.5,0.6).
- The exact distribution of $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \neq \emptyset\}$ is complicated. We use Gibbs sampler to get a sample from this distribution.

Example 3—"Fat Data"

- Let $\mathbf{X} = \mu + \sigma\mathbf{Z}$. We observe (2.0,2.1), (0.6,0.7), (0.4, 0.5), (1.4,1.5), (0.7,0.8), (0.8,0.9), (1.2,1.3), (1.2,1.3), (1.1,1.2), (1.5,1.6), (1.4,1.5), (0.4,0.5), (1.2,1.3), (0.7,0.8), (0.5,0.6).
- The exact distribution of $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \neq \emptyset\}$ is complicated. We use Gibbs sampler to get a sample from this distribution.
- Each Q is a polygon. When sampling an element of Q we take a random vertex.

Example 3—“Fat Data”

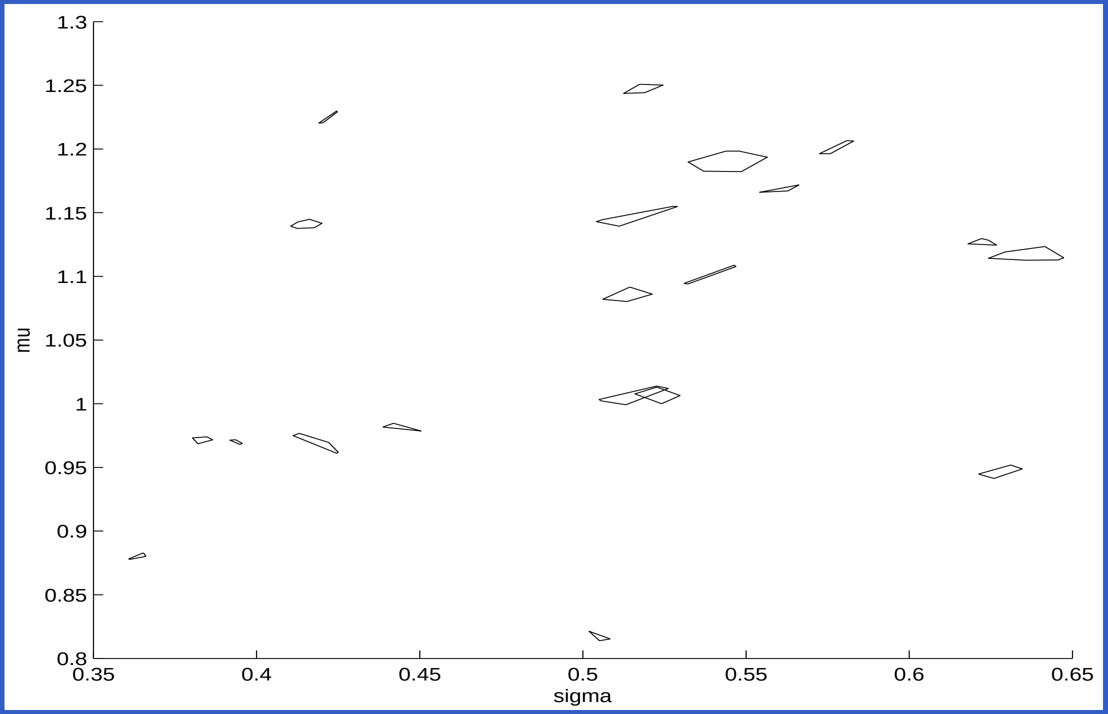
- Let $\mathbf{X} = \mu + \sigma\mathbf{Z}$. We observe (2.0,2.1), (0.6,0.7), (0.4, 0.5), (1.4,1.5), (0.7,0.8), (0.8,0.9), (1.2,1.3), (1.2,1.3), (1.1,1.2), (1.5,1.6), (1.4,1.5), (0.4,0.5), (1.2,1.3), (0.7,0.8), (0.5,0.6).
- The exact distribution of $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \neq \emptyset\}$ is complicated. We use Gibbs sampler to get a sample from this distribution.
- Each Q is a polygon. When sampling an element of Q we take a random vertex.
- Notice that this approach does not assume that the true value of $\mathbf{X}$ is uniform in the observed interval!

Example—Fat Data

A sample from $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \neq \emptyset\}$.

Example—Fat Data

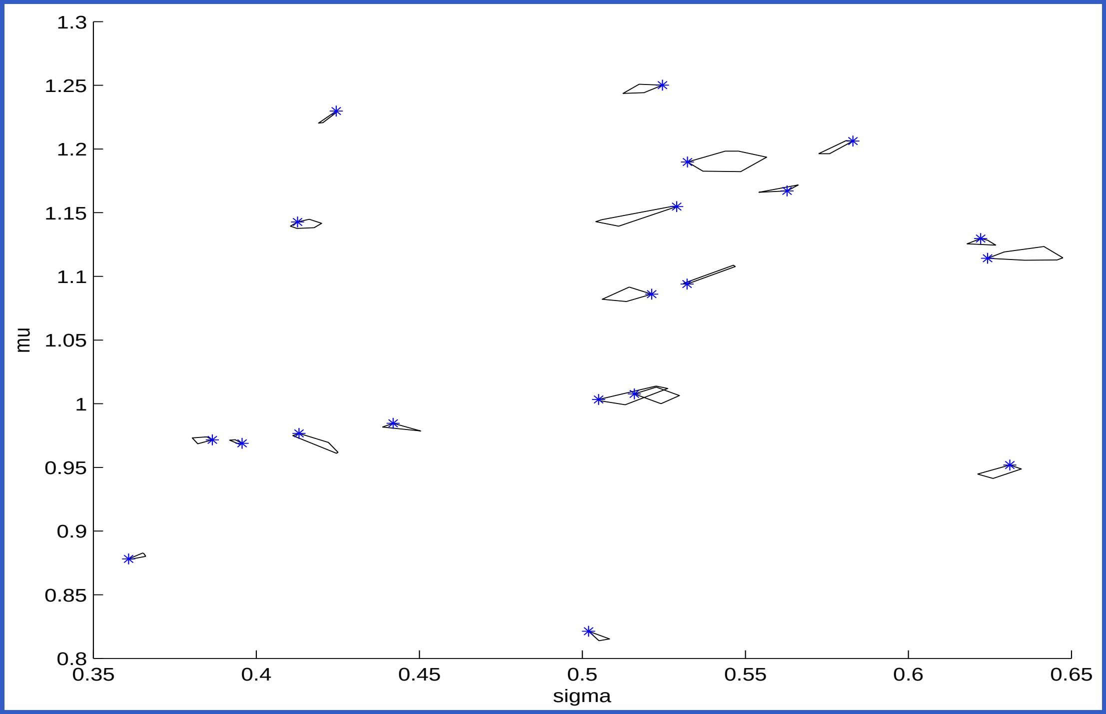
A sample from $Q((\mathbf{a}, \mathbf{b}), \mathbf{Z}^{\star}) \mid \{Q((\mathbf{a}, \mathbf{b}), \mathbf{Z}^{\star}) \neq \emptyset\}$.



20 observations.

Example—Fat Data

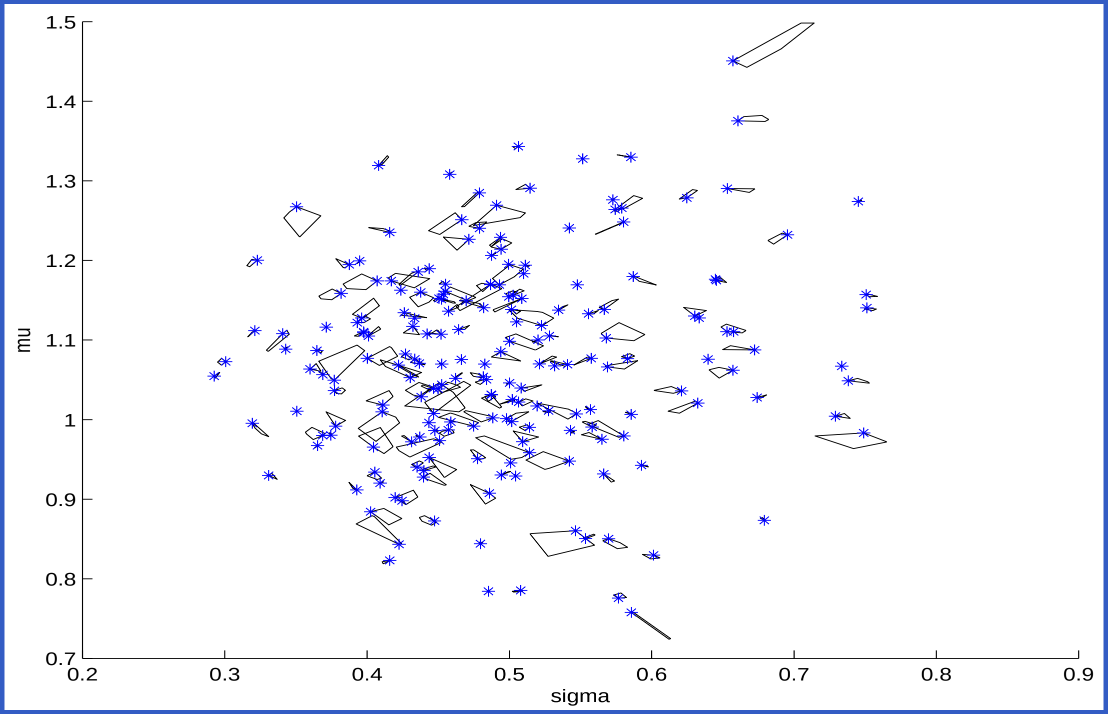
A sample from $Q((\mathbf{a}, \mathbf{b}), \mathbf{Z}^{\star}) \mid \{Q((\mathbf{a}, \mathbf{b}), \mathbf{Z}^{\star}) \neq \emptyset\}$.



20 observations, final sampled value shown.

Example—Fat Data

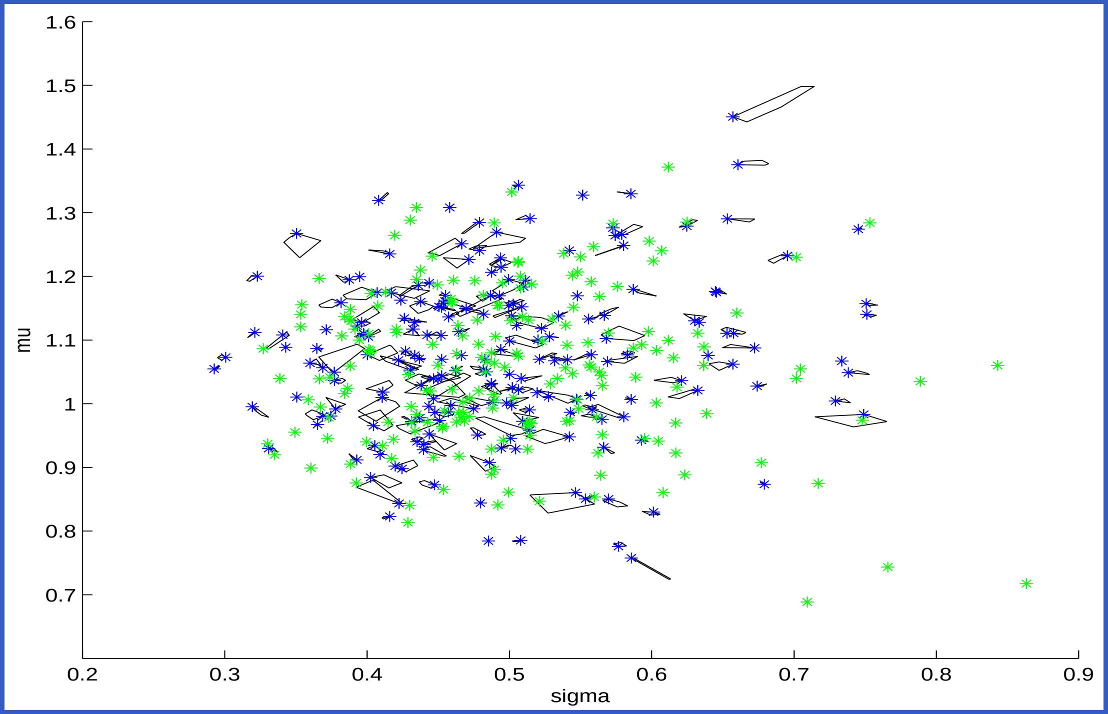
A sample from $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^{\star}) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^{\star}) \neq \emptyset\}$.



200 observations, final sampled value shown.

Example—Fat Data

A sample from $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \neq \emptyset\}$.



green sample from "usual fiducial" computed with fully known observation.

Questions

- What is the distribution of $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^{\star}) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^{\star}) \neq \emptyset\}$?
- Issues

Questions

- What is the distribution of $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \neq \emptyset\}$?
- Issues
  - Random environment: The observations X are random. But only partially observed.

Questions

- What is the distribution of $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \neq \emptyset\}$?
- Issues
  - Random environment: The observations X are random. But only partially observed.
  - The set we condition on has an extremely small probability.

Questions

- What is the distribution of $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \neq \emptyset\}$?
- Issues
  - Random environment: The observations X are random. But only partially observed.
  - The set we condition on has an extremely small probability.
  - Geometry is complicated.

Questions

- What is the distribution of $Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{Z}^\star) \neq \emptyset\}$?
- Issues
  - Random environment: The observations X are random. But only partially observed.
  - The set we condition on has an extremely small probability.
  - Geometry is complicated. It is an intersection of large number of random, dependent parallelograms. However it has typically surprisingly low number of vertexes.

What can we do?

- Let $d = (d_1, d_2) \in \mathbb{S}^2$ and define $Q_d((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star)$ the most extreme point along the direction d. (It is one of the vertexes a.s.)

What can we do?

- Let $d = (d_1, d_2) \in \mathbb{S}^2$ and define $Q_d((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star)$ the most extreme point along the direction d. (It is one of the vertexes a.s.)
- The distribution of $Q_d((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star) \mid \{Q((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star) \neq \emptyset\}$ is proportional to

$$\sum_{i<j} \frac{|c_i^{ij} - c_j^{ij}|}{\sigma^{-3}} \phi\left(\frac{c_i^{ij} - \mu}{\sigma}\right) \phi\left(\frac{c_j^{ij} - \mu}{\sigma}\right) \prod_{k \neq i,j} \left(\Phi\left(\frac{b_i - \mu}{\sigma}\right) - \Phi\left(\frac{a_i - \mu}{\sigma}\right)\right)$$

c_i^{ij} is either a_i or b_i depending on d.

What can we do?

- Let $d = (d_1, d_2) \in \mathbb{S}^2$ and define $Q_d((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star)$ the most extreme point along the direction d. (It is one of the vertexes a.s.)
- The distribution of $Q_d((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star) \mid \{Q((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star) \neq \emptyset\}$ is proportional to

$$\sum_{i<j} \frac{|c_i^{ij} - c_j^{ij}|}{\sigma^{-3}} \phi\left(\frac{c_i^{ij} - \mu}{\sigma}\right) \phi\left(\frac{c_j^{ij} - \mu}{\sigma}\right) \prod_{k \neq i,j} \left(\Phi\left(\frac{b_i - \mu}{\sigma}\right) - \Phi\left(\frac{a_i - \mu}{\sigma}\right)\right)$$

  c_i^{ij} is either a_i or b_i depending on d.
- This allows us to compute
  - Find a limit as $\mathbf{b} - \mathbf{a} \to 0$. (More on this later. 2)

What can we do?

- Let $d = (d_1, d_2) \in \mathbb{S}^2$ and define $Q_d((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star)$ the most extreme point along the direction d. (It is one of the vertexes a.s.)
- The distribution of $Q_d((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star) \mid \{Q((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star) \neq \emptyset\}$ is proportional to

$$\sum_{i<j} \frac{|c_i^{ij} - c_j^{ij}|}{\sigma^{-3}} \phi\left(\frac{c_i^{ij} - \mu}{\sigma}\right) \phi\left(\frac{c_j^{ij} - \mu}{\sigma}\right) \prod_{k \neq i,j} \left(\Phi\left(\frac{b_i - \mu}{\sigma}\right) - \Phi\left(\frac{a_i - \mu}{\sigma}\right)\right)$$

  c_i^{ij} is either a_i or b_i depending on d.
- This allows us to compute
  - Find a limit as $\mathbf{b} - \mathbf{a} \to 0$. (More on this later. 2)
  - Find a limit as $n \to \infty$. (It is consistent and asymptotically normal).

What can we do?

- Let $d = (d_1, d_2) \in \mathbb{S}^2$ and define $Q_d((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star)$ the most extreme point along the direction d. (It is one of the vertexes a.s.)
- The distribution of $Q_d((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star) \mid \{Q((\mathbf{a}, \mathbf{b}), \mathbf{Z}^\star) \neq \emptyset\}$ is proportional to

$$\sum_{i<j} \frac{|c_i^{ij} - c_j^{ij}|}{\sigma^{-3}} \phi\left(\frac{c_i^{ij} - \mu}{\sigma}\right) \phi\left(\frac{c_j^{ij} - \mu}{\sigma}\right) \prod_{k \neq i,j} \left(\Phi\left(\frac{b_i - \mu}{\sigma}\right) - \Phi\left(\frac{a_i - \mu}{\sigma}\right)\right)$$

c_i^{ij} is either a_i or b_i depending on d.

- This allows us to compute
  - Find a limit as $\mathbf{b} - \mathbf{a} \rightarrow 0$. (More on this later. 2)
  - Find a limit as $n \rightarrow \infty$. (It is consistent and asymptotically normal).
  - Would love to know $n(Q - Q_d) \mid \{Q \neq \emptyset\}$

Generalized Fiducial Recipe

- Let $\mathbb{X}$ be a random vector with a distribution indexed by a parameter $\xi \in \mathbb{R}^p$. Assume that

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

  where U has some known distribution independent of parameters, e.g, $U \sim U(0,1)$.

Generalized Fiducial Recipe

- Let $\mathbb{X}$ be a random vector with a distribution indexed by a parameter $\xi \in \mathbb{R}^p$. Assume that

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

  where U has some known distribution independent of parameters, e.g, $U \sim U(0,1)$.

- Define a set-valued function

$$Q(\mathcal{A}, u) = \{\xi : \mathbf{G}(u, \xi) \in \mathcal{A}\}.$$

  The function $Q(\mathcal{A}, U)$ is an “inverse” of the function G.

Generalized Fiducial Recipe

- Let $\mathbb{X}$ be a random vector with a distribution indexed by a parameter $\xi \in \mathbb{R}^p$. Assume that

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

where U has some known distribution independent of parameters, e.g, $U \sim U(0, 1)$.

- Define a set-valued function

$$Q(\mathcal{A}, u) = \{\xi : \mathbf{G}(u, \xi) \in \mathcal{A}\}.$$

The function $Q(\mathcal{A}, U)$ is an “inverse” of the function G.

- Assume that for any measurable S there is a random variable $V(S)$ with support S .

Generalized Fiducial Recipe

Based on $\mathbb{X} = G(U, \xi)$ define a generalized fiducial distribution as the conditional distribution of

$$V(Q(\mathcal{A}, U^\star)) \mid \{ Q(\mathcal{A}, U^\star) \neq \emptyset\}. \qquad (1)$$

Here $\mathbb{X} \in \mathcal{A}$ is the observed data and $U^\star \stackrel{\mathcal{D}}{=} U$ is a random variable independent of $\mathbb{X}$.

Generalized Fiducial Recipe

Based on $\mathbb{X} = G(U, \xi)$ define a generalized fiducial distribution as the conditional distribution of

$$V(Q(\mathcal{A}, U^\star)) \mid \{ Q(\mathcal{A}, U^\star) \neq \emptyset \}. \quad (1)$$

Here $\mathbb{X} \in \mathcal{A}$ is the observed data and $U^\star \stackrel{\mathcal{D}}{=} U$ is a random variable independent of $\mathbb{X}$.

- Let $\mathcal{R}_\xi(\mathcal{A})$ be a random variable with distribution (1). If $\theta = \pi(\xi)$ is of interest use $\mathcal{R}_\theta = \pi(\mathcal{R}_\xi)$. We will call these GFQs.

Remarks

- There are three sources of non-uniquness in (1).

Remarks

- There are three sources of non-uniquness in (1).
  - The choice of structural equation.

Remarks

- There are three sources of non-uniquness in (1).
  - The choice of structural equation.
  - The choice of $V(S)$:
    - Arises if the inverse $Q(\mathcal{A}, U^\star)$ has more then one element.

Remarks

- There are three sources of non-uniquness in (1).
  - The choice of structural equation.
  - The choice of $V(S)$:
    - Arises if the inverse $Q(\mathcal{A}, U^{\star})$ has more then one element.
  - The conditioning on $Q(\mathcal{A}, U^{\star}) \neq \emptyset\}$:
    - Arises if $P\{Q(\mathcal{A}, U^{\star}) \neq \emptyset\} = 0$.

Remarks

- There are three sources of non-uniqueness in (1).
  - The choice of structural equation.
  - The choice of $V(S)$:
    - Arises if the inverse $Q(\mathcal{A}, U^{\star})$ has more then one element.
  - The conditioning on $Q(\mathcal{A}, U^{\star}) \neq \emptyset\}$:
    - Arises if $P\{Q(\mathcal{A}, U^{\star}) \neq \emptyset\} = 0$.
    - This is caused by Borel paradox.

Remarks

- There are three sources of non-uniqueness in (1).
  - The choice of structural equation.
  - The choice of $V(S)$:
    - Arises if the inverse $Q(\mathcal{A}, U^\star)$ has more then one element.
  - The conditioning on $Q(\mathcal{A}, U^\star) \neq \emptyset\}$:
    - Arises if $P\{Q(\mathcal{A}, U^\star) \neq \emptyset\} = 0$.
    - This is caused by Borel paradox.
- Under suitable conditions the fiducial distribution leads to procedures with asymptotically correct frequentist properties.

Limit $b - a \to 0$

- Assume:
  - The model parameter ξ is p-dimensional;

Limit $b - a \to 0$

- Assume:
  - The model parameter ξ is p-dimensional;
  - the structural equation factorizes as $\mathbb{X}_0 = \mathbf{G}_0(\xi, E_0) \in \mathbb{R}^p$ and $\mathbb{X}_c = \mathbf{G}_c(\xi, E_c)$;

Limit $b - a \to 0$

- Assume:
  - The model parameter ξ is p-dimensional;
  - the structural equation factorizes as $\mathbb{X}_0 = \mathbf{G}_0(\xi, E_0) \in \mathbb{R}^p$ and $\mathbb{X}_c = \mathbf{G}_c(\xi, E_c)$;
  - $\mathbf{G}_0(\xi, \bullet\ )$ and $\mathbf{G}_c(\xi, \bullet\ )$ are one-to-one and differentiable, $\mathbf{G}_0^{-1}(\mathbf{x}_0, \xi)$ denotes the inverse; $\mathbf{G}_0(\bullet\ , \mathbf{e}_0)$ is one-to-one and differentiable.

Limit $b - a \to 0$

- Assume:
  - The model parameter ξ is p-dimensional;
  - the structural equation factorizes as $\mathbb{X}_0 = \mathbf{G}_0(\xi, E_0) \in \mathbb{R}^p$ and $\mathbb{X}_c = \mathbf{G}_c(\xi, E_c)$;
  - $\mathbf{G}_0(\xi, \bullet\ )$ and $\mathbf{G}_c(\xi, \bullet\ )$ are one-to-one and differentiable, $\mathbf{G}_0^{-1}(\mathbf{x}_0, \xi)$ denotes the inverse; $\mathbf{G}_0(\bullet\ , \mathbf{e}_0)$ is one-to-one and differentiable.
- The generalized fiducial distribution is then calculated to be

$$r(\xi|\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x}|\xi) J(\mathbf{x}, \xi)}{\int_{\Xi} f_{\mathbf{X}}(\mathbf{x}|\xi') J(\mathbf{x}, \xi')\, d\xi'}, \tag{2}$$

where $J(\mathbf{x}, \xi) = \binom{n}{p}^{-1} \sum_{\mathbf{i}} \left| \frac{\det\left(\frac{d}{d\xi} \mathbf{G}_{\mathbf{i}}^{-1}(\mathbf{x}_{\mathbf{i}}, \xi)\right)}{\det\left(\frac{d}{d\mathbf{x}_{\mathbf{i}}} \mathbf{G}_{\mathbf{i}}^{-1}(\mathbf{x}_{\mathbf{i}}, \xi)\right)} \right|$.

Why does it work asymptotically?

- The reason why generalized fiducial inference works asymptotically in frequentist sense is very similar to the reason why Bayesian inference works – Bernstein - von Mises theorem

Why does it work asymptotically?

- The reason why generalized fiducial inference works asymptotically in frequentist sense is very similar to the reason why Bayesian inference works – Bernstein - von Mises theorem
- Roughly speaking, there is a centering T such that conditionally on the data $\mathbf{X} = \mathbf{x}$ the generalized fiducial quantity $\mathcal{R}_\theta \approx N(T(\mathbf{x}), \sigma_n^2)$.

Why does it work asymptotically?

- The reason why generalized fiducial inference works asymptotically in frequentist sense is very similar to the reason why Bayesian inference works – Bernstein - von Mises theorem
- Roughly speaking, there is a centering T such that conditionally on the data $\mathbf{X} = \mathbf{x}$ the generalized fiducial quantity $\mathcal{R}_\theta \approx N(T(\mathbf{x}), \sigma_n^2)$. Moreover unconditionally $T(\mathbb{X}) \approx N(\theta, \sigma_n^2)$.

Why does it work asymptotically?

- The reason why generalized fiducial inference works asymptotically in frequentist sense is very similar to the reason why Bayesian inference works – Bernstein - von Mises theorem
- Roughly speaking, there is a centering T such that conditionally on the data $\mathbf{X} = \mathbf{x}$ the generalized fiducial quantity $\mathcal{R}_\theta \approx N(T(\mathbf{x}), \sigma_n^2)$. Moreover unconditionally $T(\mathbb{X}) \approx N(\theta, \sigma_n^2)$.
- The lower CI is approximately $(-\infty, T(\mathbf{x}) + z_\alpha \sigma_n)$. The coverage of this CI is approximately $P(\theta < T(\mathbb{X}) + z_\alpha \sigma_n) = P(-z_\alpha \sigma_n < T(\mathbb{X}) - \theta) = \alpha$.

Why does it work asymptotically?

Theorem (Hannig, 2007)**.** *Assume that* $J(x, \bullet)$ *is continuous in* θ*,* $\pi(\theta) = E_{\theta_0} J(X, \theta)$ *is finite,* $\pi(\theta_0) > 0$*, and on some neighborhood of* θ_0 $E_{\theta_0}\left(\sup_{\theta \in (\theta_0 - \delta_0, \theta_0 + \delta_0)} J(X, \theta)\right) < \infty$*. Then under regularity conditions*

$$\int_{\mathbb{R}} \left| r(\theta, \mathbf{x}) - \frac{e^{-\frac{s^2}{2/I(\theta_0)}}}{\sqrt{2\pi/I(\theta_0)}} \right| d\theta \overset{P_{\theta_0}}{\longrightarrow} 0.$$

Why does it work asymptotically?

Theorem (Hannig, 2007)**.** *Assume that* $J(x, \bullet)$ *is continuous in* θ, $\pi(\theta) = E_{\theta_0} J(X, \theta)$ *is finite,* $\pi(\theta_0) > 0$, *and on some neighborhood of* θ_0 $E_{\theta_0}\left(\sup_{\theta \in (\theta_0 - \delta_0, \theta_0 + \delta_0)} J(X, \theta)\right) < \infty$. *Then under regularity conditions*

$$\int_{\mathbb{R}} \left| r(\theta, \mathbf{x}) - \frac{e^{-\frac{s^2}{2/I(\theta_0)}}}{\sqrt{2\pi/I(\theta_0)}} \right| d\theta \overset{P_{\theta_0}}{\to} 0.$$

- The rough idea of the proof is to show that $J(x, \theta) \to \pi(\theta)$ uniformly and use Bernstein-von Mises theorem for Bayesian posterior.

Why does it work asymptotically?

Theorem (Hannig, 2007)**.** *Assume that* $J(x, \bullet)$ *is continuous in* θ*,* $\pi(\theta) = E_{\theta_0} J(X, \theta)$ *is finite,* $\pi(\theta_0) > 0$*, and on some neighborhood of* θ_0 $E_{\theta_0}\left(\sup_{\theta\in(\theta_0-\delta_0,\theta_0+\delta_0)} J(X,\theta)\right) < \infty$*. Then under regularity conditions*

$$\int_{\mathbb{R}} \left| r(\theta, \mathbf{x}) - \frac{e^{-\frac{s^2}{2/I(\theta_0)}}}{\sqrt{2\pi/I(\theta_0)}} \right| d\theta \overset{P_{\theta_0}}{\to} 0.$$

- The rough idea of the proof is to show that $J(x, \theta) \to \pi(\theta)$ uniformly and use Bernstein-von Mises theorem for Bayesian posterior. There is a technical problem caused by the fact that $\pi(\theta)$ is typically improper.

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!*"