Deep Fiducial Inference

Parts of this talk are joint work with
T. C.M Lee (UC Davis), Randy Lai (U of Maine),
Hari Iyer (NIST), Gang Li (UNC)

Summer 2019

Jan Hannig[a]

University of North Carolina at Chapel Hill

[a]NSF support acknowledged

Outline

- Introduction
- Definition
- Applications
- Conclusions

Outline

- Introduction
- Definition
- Applications
  - Deep Neural Network
- Conclusions

BFF

- Many great minds contributed to foundations of statistics in the past – Fisher, Neyman, Pierson, 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, Pierson, 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, Pierson, 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?

Bird's Eye View of Statistical Methodology

Bird's Eye View of Statistical Methodology

- Common: given X find adequate data generating mechanism

Bird's Eye View of Statistical Methodology

- Common: given X find adequate data generating mechanism
- Difference: math details, interpretation

Bird's Eye View of Statistical Methodology

- Common: given X find adequate data generating mechanism
- Difference: math details, interpretation
  - My subjective opinion: If the underlying optimization problem is the same, the methods are the same.

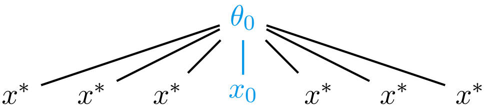
Frequentist

Frequentist

- **Modeling:** collection of distributions $\mathcal{P} = \{P_\theta\}_{\theta \in \times}$.

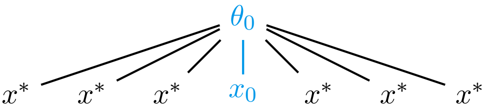
Frequentist

- **Modeling:** collection of distributions $\mathcal{P} = \{P_\theta\}_{\theta \in \times}$.
- **Replication:** parameter θ_0 fixed, data $\boldsymbol{x}$ replicated



Frequentist

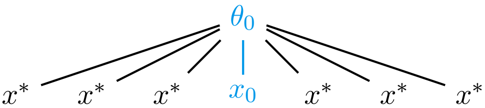
- **Modeling:** collection of distributions $\mathcal{P} = \{P_\theta\}_{\theta \in \times}$.
- **Replication:** parameter θ_0 fixed, data $\boldsymbol{x}$ replicated



- **Issues:**
  - Quality judged by averaging over unobserved data $\boldsymbol{x}^*$ (SLLN + Cournot's principle)

Frequentist

- **Modeling:** collection of distributions $\mathcal{P} = \{P_\theta\}_{\theta \in \times}$.
- **Replication:** parameter θ_0 fixed, data $\boldsymbol{x}$ replicated



- **Issues:**
  - Quality judged by averaging over unobserved data $\boldsymbol{x}^*$ (SLLN + Cournot's principle)
  - Each problem requires its own solution

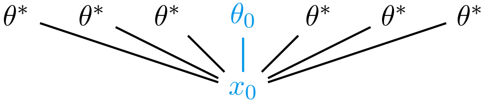
Bayesian

Bayesian

- **Modeling:** One joint distribution $f(\boldsymbol{x}|\theta) \cdot \pi(\theta)$.

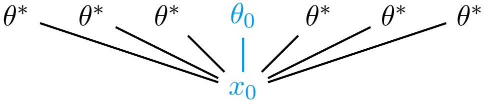
Bayesian

- **Modeling:** One joint distribution $f(\boldsymbol{x}|\theta) \cdot \pi(\theta)$.
- **Replication:** data x_0 fixed, parameter θ replicated



Bayesian

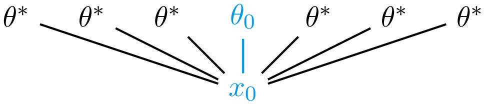
- **Modeling:** One joint distribution $f(\boldsymbol{x}|\theta) \cdot \pi(\theta)$.
- **Replication:** data x_0 fixed, parameter θ replicated



- **Issues:**
  - Averaging over unused parameters θ^* needs prior

Bayesian

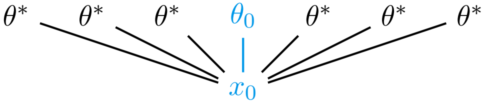
- **Modeling:** One joint distribution $f(\boldsymbol{x}|\theta) \cdot \pi(\theta)$.
- **Replication:** data x_0 fixed, parameter θ replicated



- **Issues:**
  - Averaging over unused parameters θ^* needs prior
  - Unique solution using Bayes theorem (conditional probability)

Bayesian

- **Modeling:** One joint distribution $f(\boldsymbol{x}|\theta) \cdot \pi(\theta)$.
- **Replication:** data x_0 fixed, parameter θ replicated



- **Issues:**
  - Averaging over unused parameters θ^* needs prior
  - Unique solution using Bayes theorem (conditional probability)
  - Axiomatic system for all of inference, subjective interpretation (de Finetti, Savage).

Fiducial

Fiducial

- **Modeling:** Data generating equation: $\boldsymbol{x} = G(\boldsymbol{u}, \theta)$

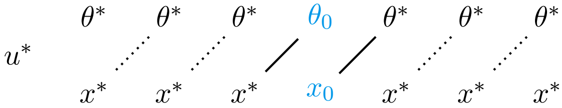
Fiducial

- **Modeling:** Data generating equation: $\boldsymbol{x} = G(\boldsymbol{u}, \theta)$
- **Replication:** data $\boldsymbol{x}$ & parameter θ linked through DGE, auxiliary variable $\boldsymbol{u}$ replicated

$$
\begin{array}{cccccccc}
 & \theta^* & \theta^* & \theta^* & \theta_0 & \theta^* & \theta^* & \theta^* \\
u_0 & \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots \\
 & x^* & x^* & x^* & x_0 & x^* & x^* & x^*
\end{array}
$$

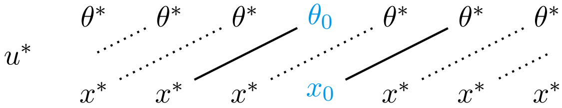
Fiducial

- **Modeling:** Data generating equation: $\boldsymbol{x} = G(\boldsymbol{u}, \theta)$
- **Replication:** data $\boldsymbol{x}$ & parameter θ linked through DGE, auxiliary variable $\boldsymbol{u}$ replicated



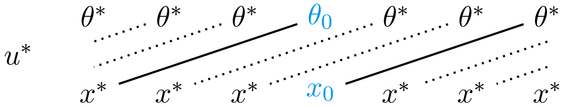
Fiducial

- **Modeling:** Data generating equation: $\boldsymbol{x} = G(\boldsymbol{u}, \theta)$
- **Replication:** data x & parameter θ linked through DGE, auxiliary variable $\boldsymbol{u}$ replicated



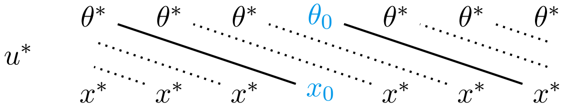
Fiducial

- **Modeling:** Data generating equation: $\boldsymbol{x} = G(\boldsymbol{u}, \theta)$
- **Replication:** data $\boldsymbol{x}$ & parameter θ linked through DGE, auxiliary variable $\boldsymbol{u}$ replicated



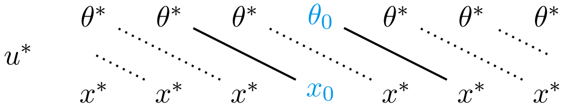
Fiducial

- **Modeling:** Data generating equation: $\boldsymbol{x} = G(\boldsymbol{u}, \theta)$
- **Replication:** data $\boldsymbol{x}$ & parameter θ linked through DGE, auxiliary variable $\boldsymbol{u}$ replicated



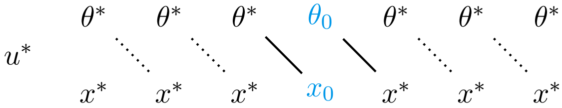
Fiducial

- **Modeling:** Data generating equation: $\boldsymbol{x} = G(\boldsymbol{u}, \theta)$
- **Replication:** data x & parameter θ linked through DGE, auxiliary variable $\boldsymbol{u}$ replicated



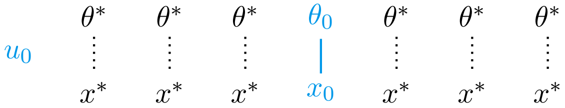
Fiducial

- **Modeling:** Data generating equation: $\boldsymbol{x} = G(\boldsymbol{u}, \theta)$
- **Replication:** data x & parameter θ linked through DGE, auxiliary variable $\boldsymbol{u}$ replicated



Fiducial

- **Modeling:** Data generating equation: $\boldsymbol{x} = G(\boldsymbol{u}, \theta)$
- **Replication:** data x & parameter θ linked through DGE, auxiliary variable $\boldsymbol{u}$ replicated



Fiducial

- **Modeling:** Data generating equation: $\boldsymbol{x} = G(\boldsymbol{u}, \theta)$
- **Replication:** data x & parameter θ linked through DGE, auxiliary variable $\boldsymbol{u}$ replicated

$$
\begin{array}{cccccccc}
 & \theta^* & \theta^* & \theta^* & \theta_0 & \theta^* & \theta^* & \theta^* \\
u_0 & \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots \\
 & x^* & x^* & x^* & x_0 & x^* & x^* & x^*
\end{array}
$$

- **Issues**
  - Fix either $\boldsymbol{x}_0$ or θ_0. Under symmetry "fiducial $\longleftrightarrow$ frequentist".

Fiducial

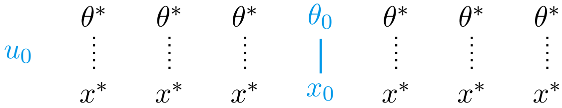
- **Modeling:** Data generating equation: $\boldsymbol{x} = G(\boldsymbol{u}, \theta)$
- **Replication:** data $\boldsymbol{x}$ & parameter θ linked through DGE, auxiliary variable $\boldsymbol{u}$ replicated

$$
\begin{array}{cccccccc}
 & \theta^* & \theta^* & \theta^* & \theta_0 & \theta^* & \theta^* & \theta^* \\
u_0 & \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots \\
 & x^* & x^* & x^* & x_0 & x^* & x^* & x^*
\end{array}
$$

- **Issues**
  - Fix either $\boldsymbol{x}_0$ or θ_0. Under symmetry "fiducial $\longleftrightarrow$ frequentist".
  - Break in symmetry: some u^* incompatible with observed $\boldsymbol{x}_0$. Still useful, frequentist properties need to be established.

Fiducial

- **Modeling:** Data generating equation: $\boldsymbol{x} = G(\boldsymbol{u}, \theta)$
- **Replication:** data $\boldsymbol{x}$ & parameter θ linked through DGE, auxiliary variable $\boldsymbol{u}$ replicated



- **Issues**
  - Fix either $\boldsymbol{x}_0$ or θ_0. Under symmetry "fiducial $\longleftrightarrow$ frequentist".
  - Break in symmetry: some u^* incompatible with observed $\boldsymbol{x}_0$. Still useful, frequentist properties need to be established.
  - Does not satisfy likelihood principle. Philosophical interpretation subject to argument

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

Outline

Comparison to likelihood

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

General Definition

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

General Definition

- Data generating equation $\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{U}, \boldsymbol{\theta})$.
  - e.g. $X_i = \theta + \sigma U_i$
- A **Generalized Fiducial Distribution** defined as a limit of

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^\star, \boldsymbol{\theta})\| \tag{1}$$

where $\boldsymbol{U}^\star$ is conditioned on

$$\{\boldsymbol{U}^\star : \ \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^\star, \boldsymbol{\theta}^*)\| \leq \varepsilon\}$$

with $\boldsymbol{\theta}$ from (1), $\varepsilon \downarrow 0$.

General Definition

- Data generating equation $\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{U}, \boldsymbol{\theta})$.
  - e.g. $X_i = \theta + \sigma U_i$
- A **Generalized Fiducial Distribution** defined as a limit of

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^\star, \boldsymbol{\theta})\| \qquad (1)$$

where $\boldsymbol{U}^\star$ is conditioned on

$$\{\boldsymbol{U}^\star : \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^\star, \boldsymbol{\theta}^*)\| \leq \varepsilon\}$$

with $\boldsymbol{\theta}$ from (1), $\varepsilon \downarrow 0$.
  - Similar to ABC; generating from prior replaced by $\min$.

General Definition

- Data generating equation $\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{U}, \boldsymbol{\theta})$.
  - e.g. $X_i = \theta + \sigma U_i$
- A **Generalized Fiducial Distribution** defined as a limit of

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^\star, \boldsymbol{\theta})\| \qquad (1)$$

where $\boldsymbol{U}^\star$ is conditioned on

$$\{\boldsymbol{U}^\star : \ \|\mathbf{x} - \boldsymbol{G}(\mathbf{U}^\star, \boldsymbol{\theta}^*)\| \leq \varepsilon\}$$

with $\boldsymbol{\theta}$ from (1), $\varepsilon \downarrow 0$.
  - Similar to ABC; generating from prior replaced by $\min$.
  - Computations?

Explicit limit (1)

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

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

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

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

Explicit limit (1)

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

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

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

- $n = p$ gives $D(A) = |\det A|$
- $\|\cdot\|_2$ gives $D(A) = (\det A^\top A)^{1/2}$
- $\|\cdot\|_\infty$ gives $D(A) = \sum_{\mathbf{i}=(i_1,\ldots,i_p)} |\det(A)_{\mathbf{i}}|$
- $\|\cdot\|_1$ gives $D(A) = \sum_{\mathbf{i}=(i_1,\ldots,i_p)} w_{\boldsymbol{i}} |\det(A)_{\mathbf{i}}|$

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

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

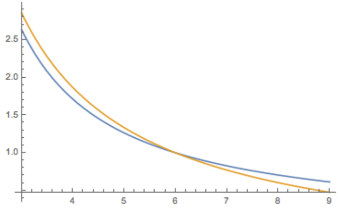
  - $= n\frac{\bar{x}(2\theta-1)-\theta^2}{\theta^2-\theta}$ for L_∞.

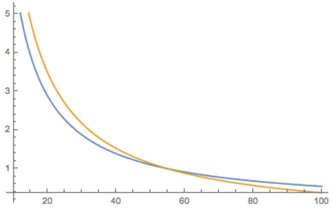
Example -- Uniform(θ, θ^2)

- 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



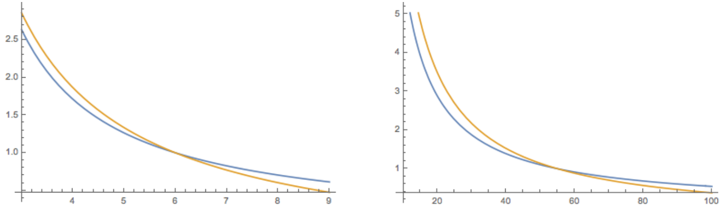


Example -- Uniform(θ, θ^2)

- 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.

Remarks

- GFD is always proper
- GFD is invariant to re-parametrizations (same as Jeffreys)
- GFD is *not* invariant to smooth transformation of the data if $n > p$
  - GFD does not satisfy likelihood principle.
- Bersntein-von Mises theorem proved in many setting

Outline

- Introduction
- Definition
- Applications
  - Deep Neural Network
- Conclusions

Outline

- Introduction
- Definition
- Applications
  - Deep Neural Network
- Conclusions

Deep Neural Network (DNN)

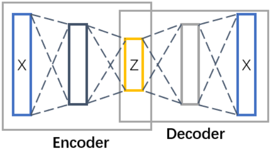
- Idea: Use deep neural network in fiducial computations

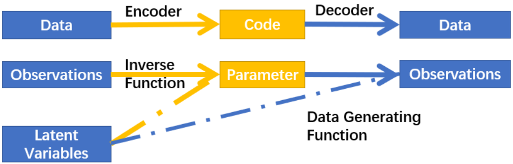
Deep Neural Network (DNN)

- Idea: Use deep neural network in fiducial computations
  - **Universal approximation theorem:** A large enough network with a linear output layer and at least one hidden layer can approximate any Borel measurable function.

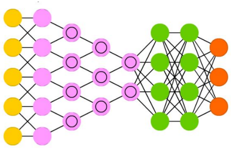
Deep Neural Network (DNN)

- Idea: Use deep neural network in fiducial computations
  - **Universal approximation theorem:** A large enough network with a linear output layer and at least one hidden layer can approximate any Borel measurable function.
  - Idea: Use Auto-encoder to approximate fiducial inverse



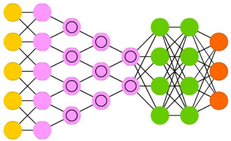


Challenges



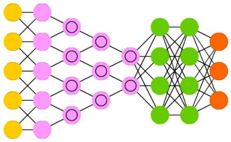
- A large number of choices
  - DNN architecture (fully connected, convolution, auto-encoder, adversarial + combination ...)

Challenges



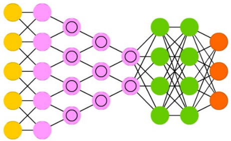
- A large number of choices
  - DNN architecture (fully connected, convolution, auto-encoder, adversarial + combination ...)
  - Number of layers, number of nodes per layers, activation function (RELU, sigmoid, softmax,...)

Challenges



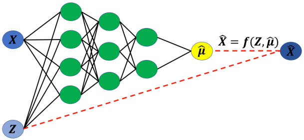
- A large number of choices
  - DNN architecture (fully connected, convolution, auto-encoder, adversarial + combination ...)
  - Number of layers, number of nodes per layers, activation function (RELU, sigmoid, softmax,...)
  - Optimization algorithm (stochastic gradient descent, Adaptive Subgradient Methods, ADAM (Kingma & Ba 2014), ...)

Challenges



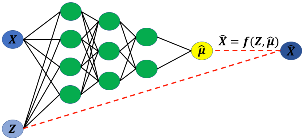
- A large number of choices
  - DNN architecture (fully connected, convolution, auto-encoder, adversarial + combination ...)
  - Number of layers, number of nodes per layers, activation function (RELU, sigmoid, softmax,...)
  - Optimization algorithm (stochastic gradient descent, Adaptive Subgradient Methods, ADAM (Kingma & Ba 2014), ...)
  - Host of other sensitivities (data generation, stopping rules, anti-over fitting measures,...)

Fiducial Auto Encoder



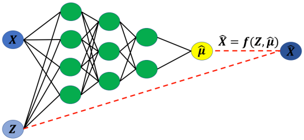
- Encoder: Fully connected layers,
- Decoder: DGE $\boldsymbol{X} = G(\boldsymbol{Z}, \boldsymbol{\mu})$

Fiducial Auto Encoder



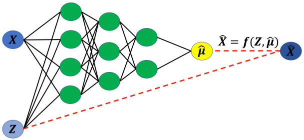
- Encoder: Fully connected layers,
- Decoder: DGE $\boldsymbol{X} = G(\boldsymbol{Z}, \boldsymbol{\mu})$
- Loss function: $L = w_1 \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2 + w_2 \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2$

Fiducial Auto Encoder



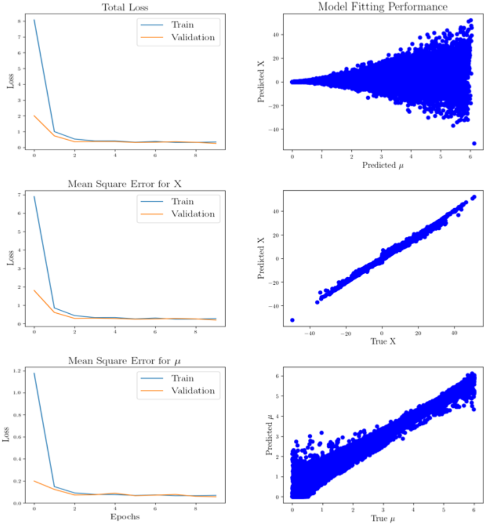
- Encoder: Fully connected layers,
- Decoder: DGE $\boldsymbol{X} = G(\boldsymbol{Z}, \boldsymbol{\mu})$
- Loss function: $L = w_1 \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2 + w_2 \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2$
- Training data: Generated from DGE with different values of $\boldsymbol{\mu}, \boldsymbol{Z}$.

Fiducial Auto Encoder



- Encoder: Fully connected layers,
- Decoder: DGE $\boldsymbol{X} = G(\boldsymbol{Z}, \boldsymbol{\mu})$
- Loss function: $L = w_1 \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2 + w_2 \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2$
- Training data: Generated from DGE with different values of $\boldsymbol{\mu}, \boldsymbol{Z}$.
- Trained encoder used for inference

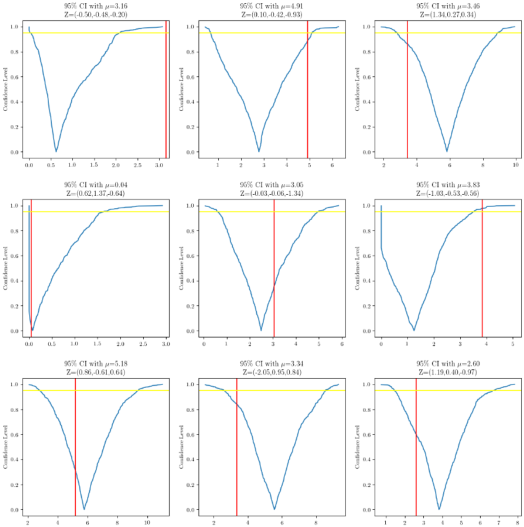
Preliminary Result - Training



- Model:
  $X_i = \mu + \mu^{q/2} Z_i$
- Network: 11 layers fully connected
- ReLU activation
- Data: Training 80,000, Validation 20,000
- Optimization: ADAM

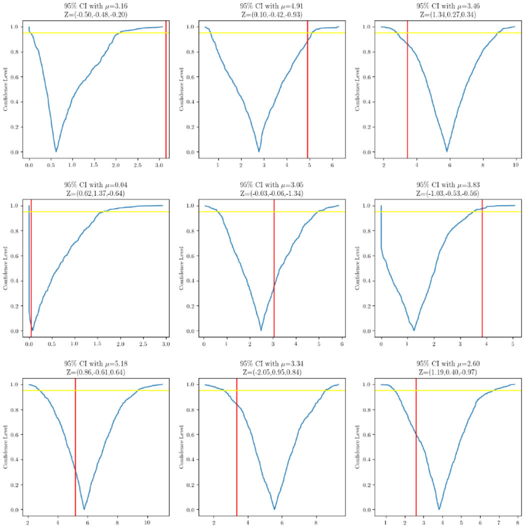
Preliminary Results - Inference

- Use encoder repeatedly
- Inputs: Observed $\boldsymbol{X}$, multiple independent $\boldsymbol{Z}^*$
- Output: Approximate fiducial sample $\boldsymbol{\mu}^*$



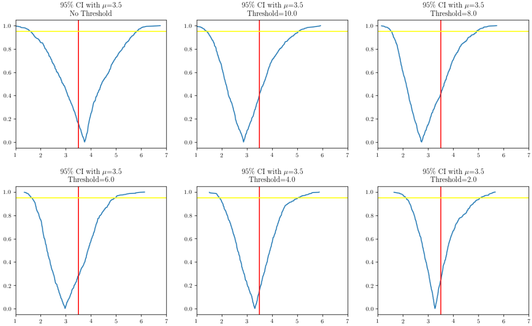
Preliminary Results - Inference

- Use encoder repeatedly
- Inputs: Observed $\boldsymbol{X}$, multiple independent $\boldsymbol{Z}^*$
- Output: Approximate fiducial sample $\boldsymbol{\mu}^*$
- Issues: conservative, biased



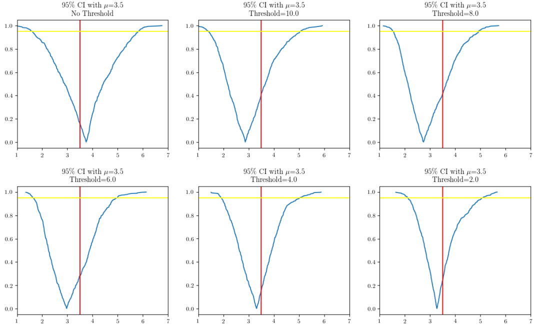
Preliminary Results - Conditioning

- Condition $\boldsymbol{Z}^*$ on $\|\boldsymbol{X}^* - \boldsymbol{X}\| \leq \epsilon$.
- Big improvement in coverage and length



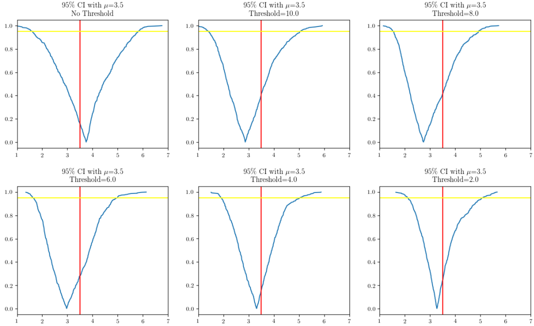
Preliminary Results - Conditioning

- Condition $\boldsymbol{Z}^*$ on $\|\boldsymbol{X}^* - \boldsymbol{X}\| \leq \epsilon$.
- Big improvement in coverage and length
- Issue: too inefficient



Preliminary Results - Conditioning

- Condition $\boldsymbol{Z}^*$ on $\|\boldsymbol{X}^* - \boldsymbol{X}\| \leq \epsilon$.
- Big improvement in coverage and length
- Issue: too inefficient
- Future work: Use GAN to generate conditional $\boldsymbol{Z}^*$.



Outline

- Introduction
- Definition
- Applications
  - Deep Neural Network
- Conclusions

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 one day soon 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 one day soon gain similar trust in fiducial inspired approaches.

Thank you!