Approximate Fiducial Computation and Deep Fiducial Inference

THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL

Gang Li* (franklee@live.unc.edu) and Jan Hannig

Department of Statistics and Operations Research, UNC at Chapel Hill, NC 27599

BACKGROUND

Generalized Fiducial Distribution

- **Data Generating Function (Data Generating Equation):**

$$X = f(Z, \mu)$$

  - X represents the data we observed;
  - The random latent variable Z has **known** distribution F_0.
  - The deterministic function f is known.
  - The parameter μ is **fixed**;
- **Generalized Fiducial Distribution (GFD):** A probability measure on the parameter space defined as

$$\lim_{\epsilon \to 0} \left[\arg\min_{\mu^*} \|x - f(Z^*, \mu^*)\| \mid \min_{\mu^*} \|x - f(Z^*, \mu^*)\| \le \epsilon \right]$$

User-friendly Formula

Theorem ([Hannig et al., 2016])

Under mild condition, the limiting distribution above has a density

$$f(\boldsymbol{\mu}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\mu}) J(\boldsymbol{x}, \boldsymbol{\mu})}{\int f(\boldsymbol{x}|\boldsymbol{\mu}') J(\boldsymbol{x}, \boldsymbol{\mu}') \, d\boldsymbol{\mu}'}$$

where $f(\boldsymbol{x}|\boldsymbol{\mu})$ is the likelihood function and

$$J(x, \mu) = D\left(\nabla_\mu f(z, \mu)|_{z=f^{-1}(x,\mu)} \right)$$

with $D(A) = (\det A'A)^{\frac{1}{2}}$.

- **Forward Solution:** Standard MCMC-type sampling techniques have already been implemented [Hannig et al., 2016]. **However, sometimes the generalized fiducial density can be hard to compute.**

Backward Solution

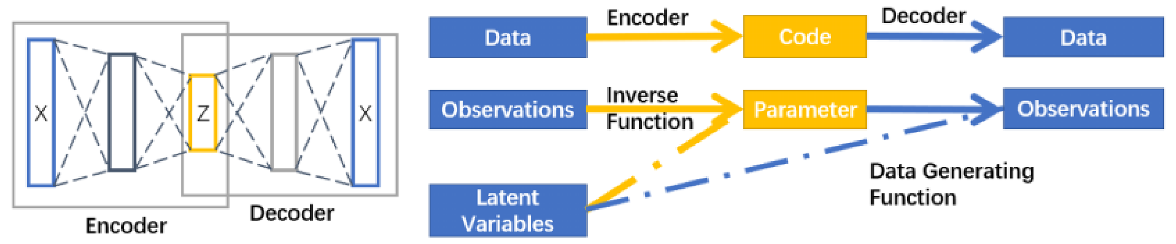
- **Inverse Function** If the data generating function is **monotone** in μ, then there exists a unique inverse function: $\mu = g(X, Z)$.
- **Backward Solution** If the **exact form** of the inverse function exists, we can sample from the approximate generalized fiducial distribution through the approximate fiducial computation (AFC) algorithm;
- AFC $\Longrightarrow$ GFS $\Longrightarrow$ Generalized Fiducial Distribution (GFD).
- With GFD, one can construct most common inference procedures (point estimates, confidence intervals and so on).

METHODS

Approximate Fiducial Computation

Output: Generalized Fiducial Samples(GFS)
Initialization: $itr = 0$; $GFS = \emptyset$;
while $(itr < max\_itr)$ *and* $(\textbf{card}(GFS) < N)$;
do
  Sample z from F_0;
  $\hat{\mu} = g(x, z)$;
  $\hat{x} = f(z, \hat{\mu})$;
  if $dist(x, \hat{x}) < Threshold$ **then**
    Accept $\hat{\mu}$;
  else
    Reject $\hat{\mu}$;
  end
  $itr = itr + 1$;
end

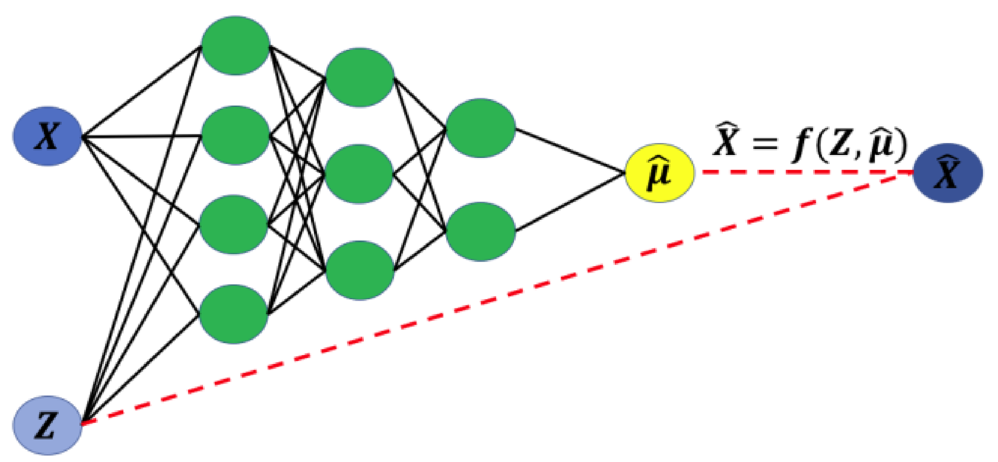
Algorithm 1: Approximate Fiducial Computation(AFC)



- **Punchline: The inverse function plays a similar role as the encoder.**
- Fiducial Autoencoder (FAE): A convenient and efficient computation tool, FAE, is implemented to generate the fiducial distribution without knowing the exact form the density.

Deep Fiducial Inference

- **Fiducial Autoencoder (FAE)**

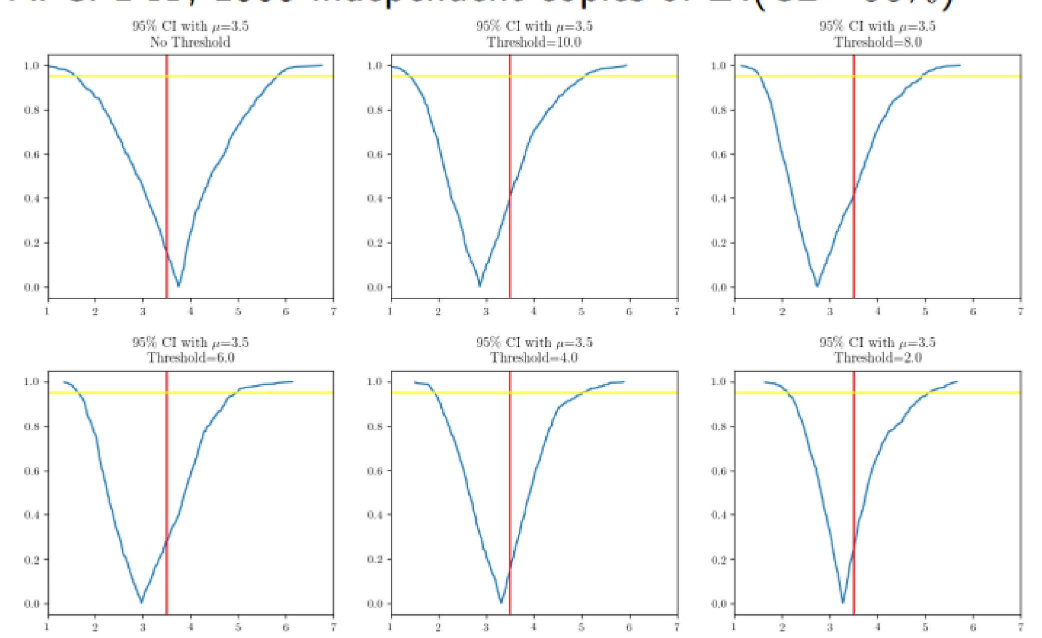


- **Loss Function:** $L = w_1\|x - \hat{x}\|^2 + w_2\|\mu - \hat{\mu}\|^2$
- **Convenience of DFI: Finite observations? Not a problem!**
- For training the FAE and approximating the inverse function, we could **simulate infinite** pairs of X and Z.

SIMULATION I

AFC with Decreasing Thresholds

- $\mathbf{x} = \mu \times \mathbf{1} + \mu^{\frac{q}{2}} \times \mathbf{z}$ where $\mathbf{x} \in \Re^m$, $\mu \in \Re$, and $\mathbf{z} \in N_m(0, I)$, $q = 3$, $m = 3$.
- With AFC: 1 X, 1000 independent copies of Z.(CL$= 95\%$)



- Both the point estimation and confidence interval estimation improves. Note decreasing the threshold might not always help.

- **Efficiency and Accuracy Trade-off:**
- Selection of Threshold:
- If the threshold is too big $\Longrightarrow$ biased samples;
- If the threshold is too small, $\Longrightarrow$ very difficult to generate enough samples.
- **Efficiency and Accuracy Trade-off:** In practice, we could use one random batch of samples to estimate the distribution of $dist(x, \hat{x})$, and select the threshold.

Nonlinear Data Generating Equation

200 observations, 1000 generalized fiducial samples, and confidence level $= 90\%$

True μ	Coverage	Expected CI Length	Expected Mean	Expected Median
1	0.985	2.03	1.07	0.9
2	0.905	3.5	2.64	2.49
3	0.865	4.25	3.85	3.81
4	0.87	4.28	4.48	4.45

Table 1: Inference Performance **without AFC**.

200 observations, 1000 generalized fiducial samples, and confidence level $= 90\%$

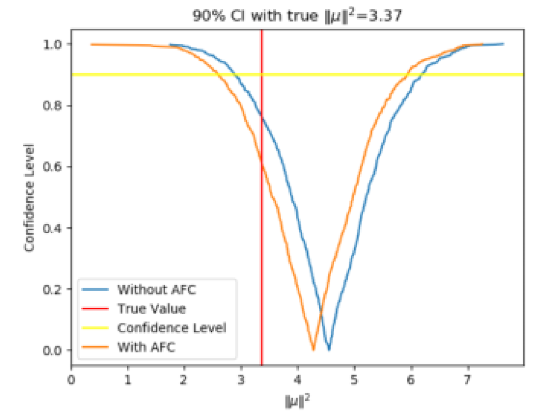
True μ	Coverage	Expected CI Length	Expected Mean	Expected Median
1	0.95	3.1	1.44	1.06
2	0.95	4.07	2.55	2.18
3	0.97	3.43	3.22	2.99
4	0.94	3.3	3.98	3.89

Table 2: Inference Performance **with AFC**.

SIMULATION II

Many Means

- **Many Means:** $\mathbf{x} = \mu + \mathbf{z}$, with $\mathbf{x}, \mu, \mathbf{z} \in \Re^m$ and $m = 3$
- The parameter of interests is the square of the l-2 norm of μ, $\|\mu\|^2$, instead of μ itself.
- Since $\hat{\mathbf{x}} = f(\hat{\mu}, \mathbf{z}^*) = \hat{\mu} + \mathbf{z}^* = g(\mathbf{x}, \mathbf{z}^*) + \mathbf{z}^* = \mathbf{x} - \mathbf{z}^* + \mathbf{z}^* = \mathbf{x}$, We proposed the following estimation $\tilde{\mu} = \|\hat{\mu}\| * (\mathbf{x}/\|\mathbf{x}\|)$ with $\hat{\mu} = \mathbf{x} - \mathbf{z}^*$.
- Confidence Curves with and without AFC.(CL$= 90\%$)



200 observations, 1000 generalized fiducial samples, and confidence level $= 90\%$

If AFC	Coverage	Expected CI Length	Expected Mean	Expected Median
No	0.85	3.12	3.91	3.91
Yes	0.95	3.29	3.65	3.64

Table 3: Inference Performance with and without AFC. (True $\|\mu\|^2 = 3.37$)

As shown in Table 3, the empirical coverage increase 10%. In addition, both the expected median and the expected mean are more accurate.

CONCLUSIONS

Conclusions

- In summary, we introduce AFC and provide a backward solution for generalized fiducial inference.
- We further design FAE for the circumstance in which the analytical form of the inverse function is not available.
- The universal approximation theorem provides theoretical guarantees for the approximation performance.
- Our simulation validates our approach.
- Further research: real data applications.

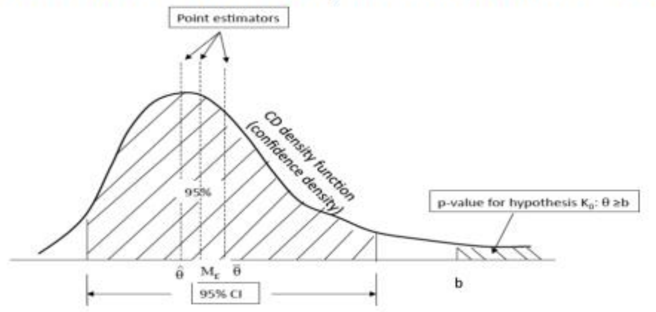
AFC and DFI Supplementary Poster

Gang Li* (franklee@live.unc.edu) and Jan Hannig
Department of Statistics and Operations Research, UNC at Chapel Hill, NC 27599
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

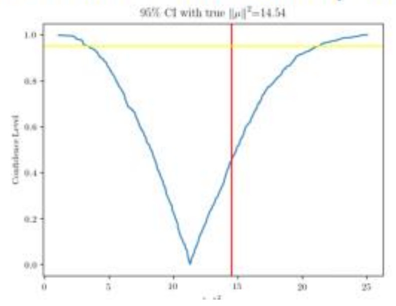
BACKGROUND

Distribution Estimator

- **Distribution Estimator**
  Point Estimator; Interval Estimator; Distribution Estimator



- **Confidence Curve(CC)** CC is a great visualization tool for GFD. CC is defined as $\mathrm{CV}(\mu) = 2|F(\mu|\mathbf{x}) - 0.5|$. CC shows confidence intervals at all significance levels stacked up on each other.



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

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

METHODS

Deep Fiducial Inference

- **Challenge of Backward** The inverse function might not have an analytical form or might be difficult to calculate even if it exists.
- **Deep Fiducial Inference** Use deep neural network to approximate the inverse function and then apply AFC:

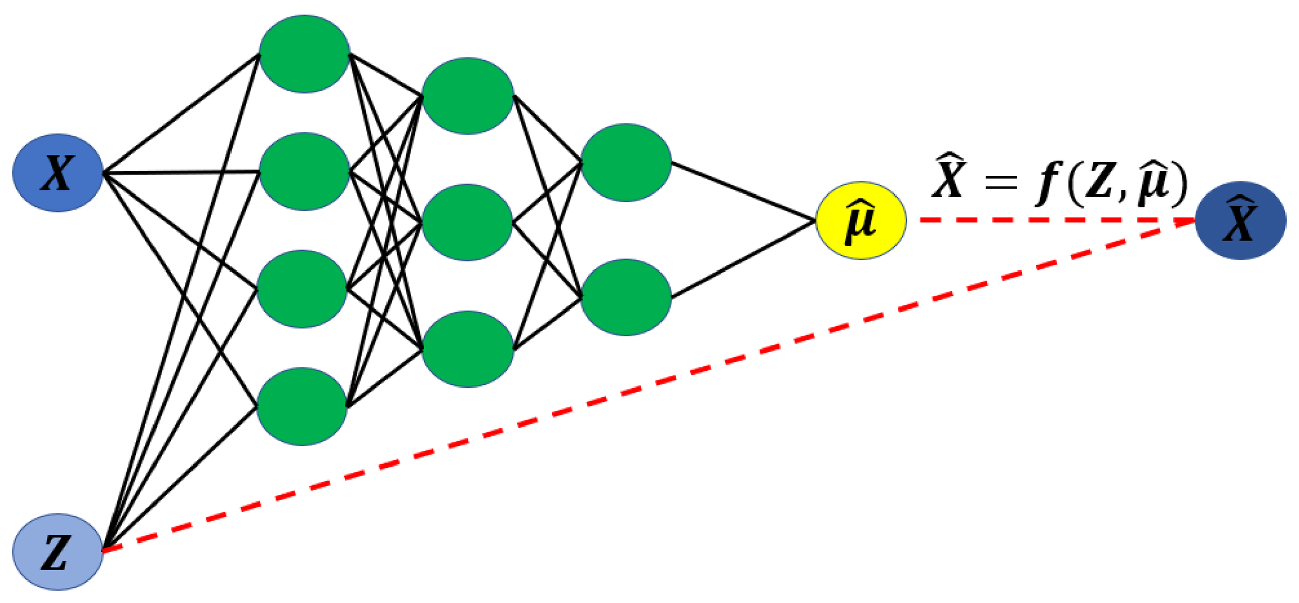
$$\hat{\mu} = \hat{g}_{NN}(X, Z)$$

- Reason for Using NN:

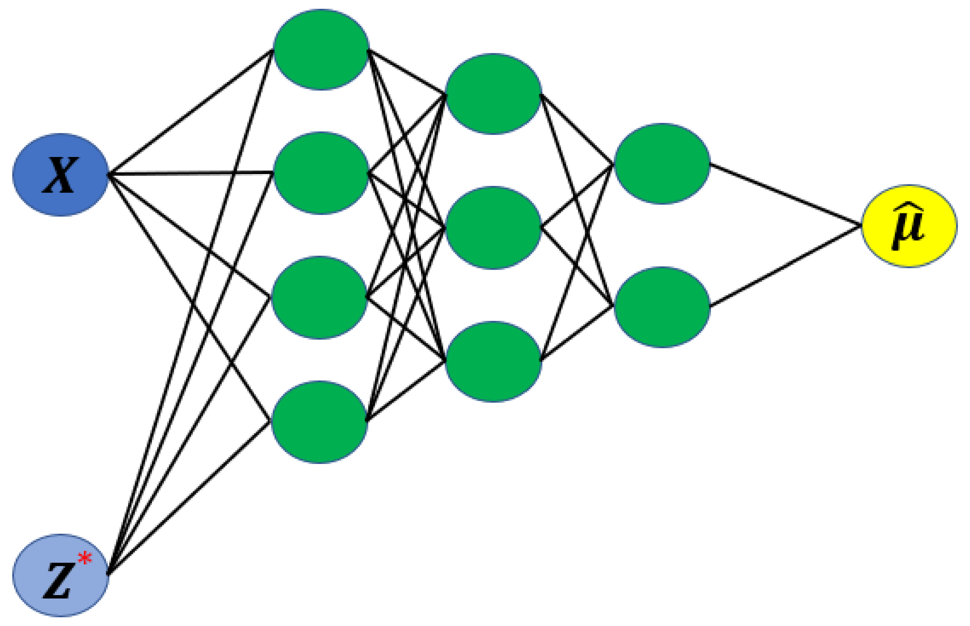
Theorem (Universal Approximation Theorem [Hornik et al., 1989])

A feedforward network with a linear output layer and at least one hidden layer with any "squashing" activation functions can approximate any Borel measurable function, provided that the network is given enough hidden units.

Fiducial Autoencoder Training



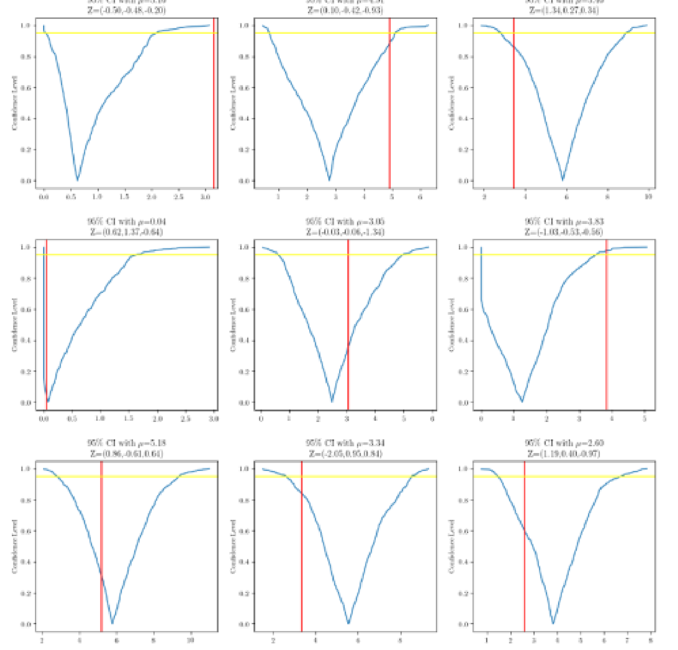
Fiducial Autoencoder Prediction



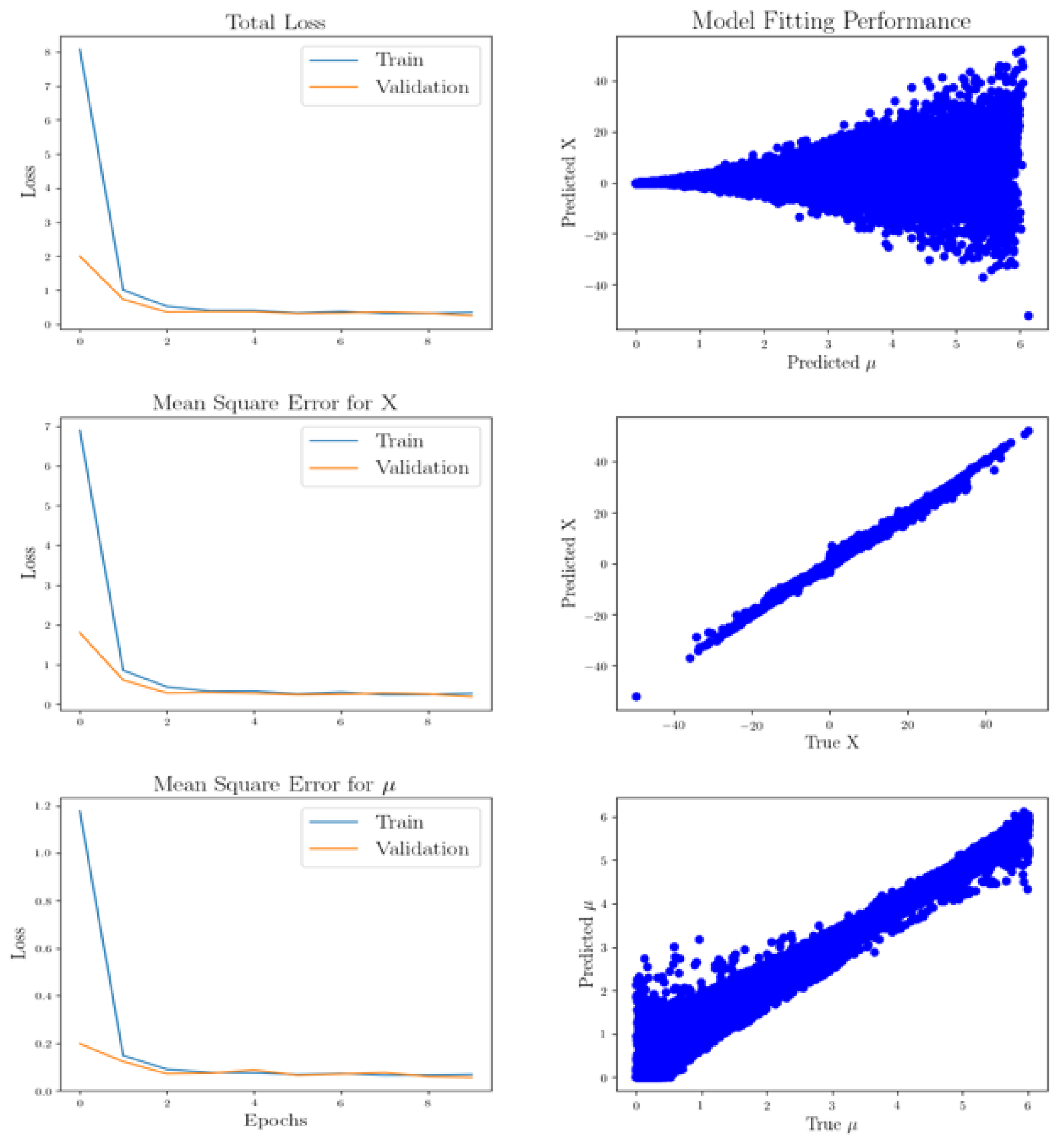
RESULTS

Nonlinear Data Generating Function (NDGF)

- **NDGF:** $\mathbf{x} = \mu \times \mathbf{1} + \mu^{\frac{q}{2}} \times \mathbf{z}$ where $\mathbf{x} \in \Re^m$, $\mu \in \Re$, and $\mathbf{z} \in N_m(0, I)$, $q = 3$, $m = 3$.
- **Without AFC:** 1 X, 1000 independent copies of Z.(CL$= 95\%$)

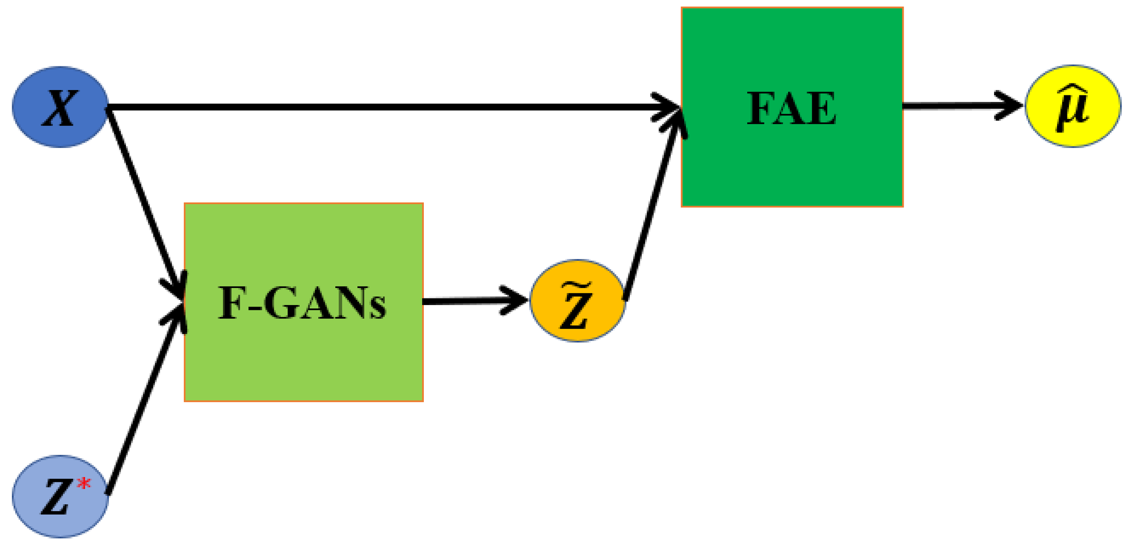


- Only 7 out of 9 are covered by the fiducial interval.
- the confidence interval is very wide.

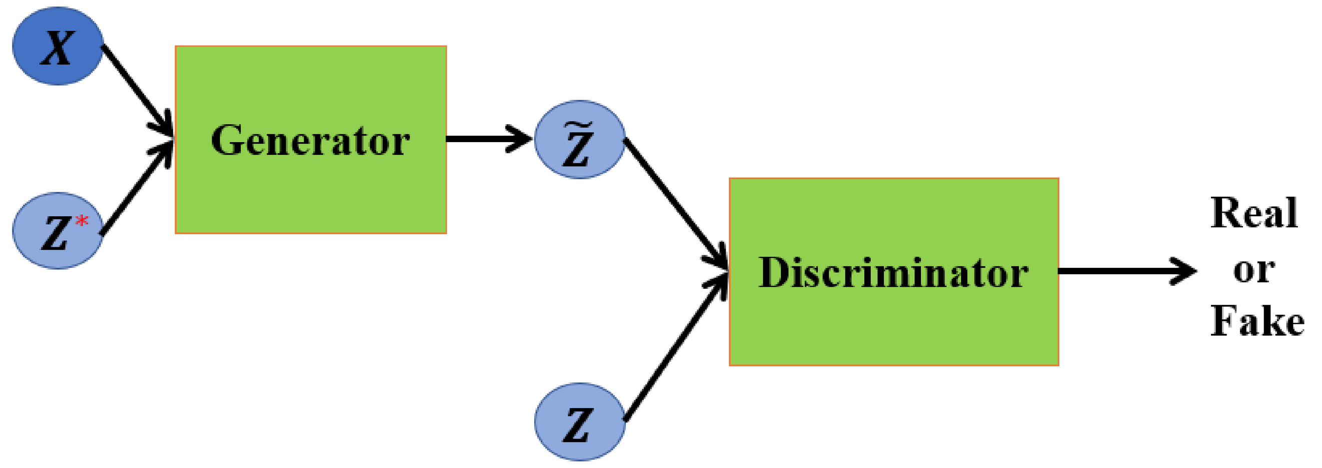


FUTURE WORK

DFI 2.0 = F-GANs + FAE



F-GANs (Fiducial General Adversarial Networks)



REFERENCES

Birnbaum, A. (1961).
On the foundations of statistical inference: binary experiments.
The Annals of Mathematical Statistics, pages 414–435.

Hannig, J., Iyer, H., Lai, R. C., and Lee, T. C. (2016).
Generalized fiducial inference: A review and new results.
Journal of the American Statistical Association, 111(515):1346–1361.

Hornik, K., Stinchcombe, M., and White, H. (1989).
Multilayer feedforward networks are universal approximators.
Neural Networks, 2(5):359–366.

Kingma, D. P. and Ba, J. (2014).
Adam: A method for stochastic optimization.
arXiv preprint arXiv:1412.6980.

Nair, V. and Hinton, G. E. (2010).
Rectified linear units improve restricted boltzmann machines.
In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

Xie, M.-G. and Singh, K. (2013).
Confidence distribution, the frequentist distribution estimator of a parameter: A review.
International Statistical Review, 81(1):3–39.