Towards Generalized Fiducial Inference for Finite Mixtures

Derek S. Young

Dr. Bing Zhang Department of Statistics
University of Kentucky

Joint work with Jan Hannig (UNC-Chapel Hill)

WGMBC 2023, Pittsburgh, PA
July 20^{th}, 2023

College of Arts and Sciences
Dr. Bing Zhang Department of Statistics

Outline of Topics

Outline of Topics

Some History

- Origins of fiducial inference can be traced back to Fisher (1922), who introduced a fiducial distribution for a parameter – in place of the Bayesian posterior – for interval estimation of said parameter
- **Single-parameter families of distributions**: Fiducial intervals coincide with classical confidence intervals
- **Multi-parameter families of distributions**: Fiducial approach yields confidence sets with frequentist coverage probabilities close to the nominal level, but are not exact in the repeated sampling frequentist sense
- Mid-20th century: Prominent statisticians penned many critical discussions about the fiducial argument
- Late-20th century: Infrequent publications on the topic, with it seemingly becoming a topic of mere historical interest
- Early-21st century: A revival of interest in modern modifications of fiducial inference
- See Hannig et al. (2016) for a contemporary review on the topic, including key references traversing the timeline stated above

Generalized Fiducial Inference (Hannig, 2009)

- **Generalized fiducial inference (GFI)** aims to define a distribution for parameters of interest that contains all the information from data
  - The paradigm carefully uses an inverse of a deterministic data-generating equation without the use of Bayes' theorem
- Inference for the parameters can therefore be made from this **(generalized) fiducial distribution**, which can further be interpreted as a posterior distribution without assuming a prior distribution (Efron, 1998)
- The random variable having a derived fiducial distribution is called a **generalized fiducial quantity (GFQ)**
- The tenet of the GFI framework is to switch the role of the parameters and the data
- Unfortunately, there is typically no unique way to define a fiducial distribution

Brief Mathematical Setup

- Suppose the data $\mathbf{X}$ are generated through the structural equation $\mathbf{X} = G(\boldsymbol{\xi}, U)$
  - $\boldsymbol{\xi} \in \Xi$ is a vector of parameters
  - U is some random variable with a known distribution independent of $\boldsymbol{\xi}$
  - The structural equation can be regarded as a data generation process where the noise process $U = u$ and the signal $\boldsymbol{\xi}$ will produce observed data $\mathbf{X} = \mathbf{x}$
- Hence, the distribution of $\mathbf{X}$ can be determined via the structural equation given a fixed parameter $\boldsymbol{\xi}$ and the distribution U
- After the data $\mathbf{X}$ are observed, switch the position of the data and parameters by solving the structural equation (conditioned on the existence of the solution)
- Thus, we get $\boldsymbol{\xi} = Q(\mathbf{X}, U)$, where $Q(\mathbf{X}, U)$ is the inverse function used to define the following generalized fiducial distribution on Ξ: $V(Q(\mathbf{x}, U^*))|\{Q(\mathbf{x}, U^*) \neq \emptyset\}$, where U^* is an independent copy of U
- A random element generated from this fiducial distribution, say $\mathcal{R}_{\boldsymbol{\xi}}(\mathbf{x})$, is a GFQ

Outline of Topics

Model and Notation

- Hannig (2009) considered the generalized fiducial distribution for the parameters of a mixture of two normal distributions
- Let $X_1, \ldots, X_n$ be independent random variables drawn from a classic five-parameter, two-component GMM:

$$(1-\pi)\mathcal{N}(\mu_1, \sigma_1^2) + \pi\mathcal{N}(\mu_2, \sigma_2^2)$$

- Assumptions:
  - $\mu_1 < \mu_2$ (identifiability constraint)
  - We observe at least two data points from each distribution
- Goal: Find the generalized fiducial distribution of $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \pi)^{\mathrm{T}}$

Structural Equations

- We can write a set of structural equations for $X_1, \ldots, X_n$ as

$$X_i = (\mu_1 + \sigma_1 Z_i)\, \mathtt{I}_{\{(0,\pi)\}}(U_i) + (\mu_2 + \sigma_2 Z_i)\, \mathtt{I}_{\{(\pi,1)\}}(U_i), \quad i = 1, \ldots, n,$$

  where $Z_i \overset{iid}{\sim} \mathcal{N}(0,1)$ and $U_i \overset{iid}{\sim} \mathcal{U}(0,1)$
- When finding the inverse (set-valued) function Q, this inversion will be stratified based on the possible assignment of the observed x_i to one of the two components
- The Q function, which is omitted for brevity (see p. 528 of Hannig (2009)), is an extension to the framework for finding the generalized fiducial distribution of (μ, σ^2) in the $\mathcal{N}(\mu, \sigma^2)$ setting
- The sums in the generalized fiducial distribution have a total of $2^n - 2 - 2n - n(n-1)$ terms, so we are unable to get a closed-form generalized fiducial density
- Turn to a Metropolis-Hastings algorithm to simulate observations from the derived generalized fiducial distribution to perform inference

k-Component GMMs

- Now let $X_1, \ldots, X_n$ be independent random variables drawn from a k-component ($k > 2$) GMM:

$$X_i \sim \mathcal{N}(\mu_j, \sigma_j^2) \quad \text{with probability } \mathrm{P}\{W_i = j\} = \pi_j,$$

  where the W_i is a membership variable

- Assumptions:
  - $\mu_1 < \mu_2 < \cdots < \mu_k$ (identifiability constraint)
  - We observe at least two data points from each distribution
- The number of occurrences of the outcome j among $W_1, \ldots, W_n$ is denoted by n_j; i.e., $\sum_{i=1}^{n} \mathtt{I}\{W_i = j\} = n_j$, such that $\sum_{j=1}^{k} n_j = n$
- We can apply the recipe used in Hannig (2009) for the $k = 2$ setting to the above $k > 2$ setting

GFQ for π_j

- The W_i, $i = 1, \ldots, n$, can be treated as outcomes from the following data-generating equation:

$$W_i = \sum_{j=0}^{k} \mathtt{I}\left\{ U_i \in \left[\sum_{l=0}^{j} \pi_l, 1 \right] \right\},$$

  where $U_i \overset{iid}{\sim} \mathcal{U}(0,1)$ and $\pi_0 = 0$

- A GFQ for π_j can be expressed as

$$\mathcal{R}_{\pi_j} = \begin{cases} U_{(r_j)} + D_j[U_{(r_j+1)} - U_{(r_j)}] & j = 1; \\ U_{(r_j)} + D_j[U_{(r_j+1)} - U_{(r_j)}] - \mathcal{R}_{\pi_{j-1}} & j = 2, \ldots, k-1; \\ 1 - \sum_{l=1}^{k-1} \mathcal{R}_{\pi_l} & j = k, \end{cases}$$

  where $U_{(1)}, \ldots, U_{(n)}$ are the order statistics of $U_1, \ldots, U_n$, $r_j = \sum_{l=1}^{j} n_l$, and $D_j \overset{iid}{\sim} \mathcal{U}(0,1)$

- In the formula for $\mathcal{R}_{\pi_j}$, we set $U_{(0)} = 0$ and $U_{(n+1)} = 1$

GFQs for μ_j and σ_j^2

- We extend the set of structural equations for $X_1, \ldots, X_n$ used in the two-component setting to the k-component setting
- A GFQ for σ_j^2 can be expressed as

$$\mathcal{R}_{\sigma_j^2} = \frac{(n_j - 1)s_j^2}{V_j},$$

where s_j^2 denotes the sample variance and $V_j \sim \chi^2_{n_j - 1}$

- A GFQ for μ_j can be expressed as

$$\mathcal{R}_{\mu_j} = \bar{x}_j - Z_j \sqrt{\frac{\mathcal{R}_{\sigma_j^2}}{n_j}},$$

where $\bar{x}_j$ denotes the sample mean and $Z_j \sim \mathcal{N}(0, 1)$

Sketch of MCMC Sampler

1. Initialize the sampler by determining an arbitrary assignment to the k components, say, $\mathbf{w}^{(0)} = (w_1^{(0)}, \ldots, w_n^{(0)})^{\mathrm{T}}$
2. Generate a proposal configuration by taking the previous assignment, randomly choose one data point, and switch it to another component (accept/reject based on usual Metropolis-Hastings rule)
3. Based on the current assignment, $\mathbf{w}^{(t)}$, generate realizations of $\mathcal{R}_{\mu_j}$, $\mathcal{R}_{\sigma_j^2}$, and $\mathcal{R}_{\pi_j}$, $j = 1, \ldots, k$
4. The stationary distribution of the assignment-valued Markov chain is the generalized fiducial distribution of the assignment

Using the Results

- Since a generalized fiducial distribution provides us with a distribution on the parameter space, its use is similar to the practical use of a Bayesian posterior
- After a burn-in period, we can take, for example, the mean to get a point estimator of the full parameter vector $\boldsymbol{\xi} = (\mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2, \pi_1, \ldots, \pi_{k-1})^{\mathrm{T}}$
  - Posterior membership probabilities can then be calculated for doing model-based clustering
- We can find $\mathcal{C}(\mathbf{x})$ with fiducial probability $\mathrm{P}\left\{\mathcal{R}_{\boldsymbol{\xi}}(\mathbf{x}) \in \mathcal{C}(\mathbf{x})\right\} = 1 - \alpha$ to get approximate $100 \times (1 - \alpha)\%$ fiducial confidence sets
  - These confidence sets, though not exact, often have very good coverages and expected length properties in small sample simulations, but can be exact asymptotically

Example: Simulated Data

- $k = 3$ components
- $\boldsymbol{\xi} = (\mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2, \pi_1, \pi_2)^{\mathrm{T}} = (0, 6, 12, 1, 1, 1, 0.50, 0.25)^{\mathrm{T}}$
- $n = 100$
- Generated $M = 5000$ fiducial samples after dropping 5000 for burn-in
- Computed point estimates based on the fiducial approach and compared with the maximum likelihood solutions using EM
- Code is available at my `GitHub` repo: https://github.com/dsy109/Supplemental/blob/main/WGMBC/MixNormFid.R

Example: Simulated Data (ctd.)

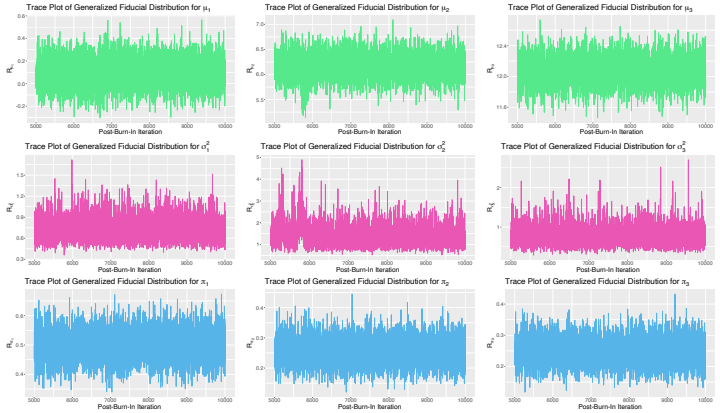


Figure 1: Trace plots of the generalized fiducial distributions

Example: Simulated Data (ctd.)

Parameter	Fiducial	EM
μ_1	0.0980	0.1004
μ_2	6.1604	6.1657
μ_3	12.0699	12.0693
σ_1^2	0.7236	0.6774
σ_2^2	1.3740	1.1749
σ_3^2	0.7533	0.6626
π_1	0.4998	0.5000
π_2	0.2530	0.2500

Table 1: Fiducial and EM estimates of $\boldsymbol{\xi}$ for the simulated data

Example: 1872 Hidalgo Stamp Data

- Analyzed the famous 1872 Hidalgo stamp data ($n = 485$) assuming $k = 4$ components (Izenman & Sommer, 1988)
- Generated $M = 50000$ fiducial samples after dropping 50000 for burn-in
- Trace plots indicate convergence

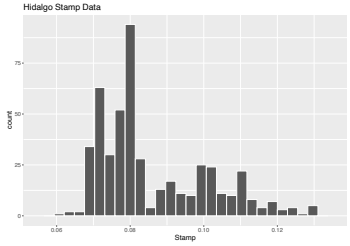


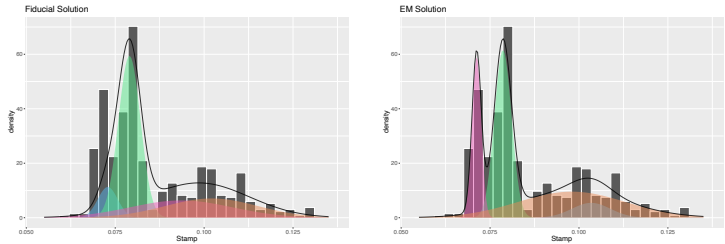
Figure 2: Histogram of Hidalgo stamp data

Example: 1872 Hidalgo Stamp Data (ctd.)

Parameter	Fiducial	EM
μ_1	0.0729	0.0712
μ_2	0.0790	0.0786
μ_3	0.0935	0.0980
μ_4	0.1021	0.1034
σ_1	0.0028	0.0013
σ_2	0.0031	0.0024
σ_3	0.0143	0.0151
σ_4	0.0131	0.0054
π_1	0.0789	0.1926
π_2	0.4620	0.3722
π_3	0.2278	0.3613

Table 2: Fiducial and EM estimates of $\boldsymbol{\xi}$ for the Hidalgo stamp data

Example: 1872 Hidalgo Stamp Data (ctd.)



(a) (b)

Figure 3: (a) Fiducial fits and (b) EM fits for GMM with $k = 4$ components

Outline of Topics

Model Selection

- Consider a finite collection of models $\mathcal{M}$
- Data-generating equation is $\mathbf{X} = G(M, \boldsymbol{\xi}_M, U)$, $M \in \mathcal{M}$, $\boldsymbol{\xi}_M \in \boldsymbol{\Xi}_M$, where M is the model considered and $\boldsymbol{\xi}_M$ are the parameters associated with model M
- Similar to maximum likelihood estimation, GFI tends to favor models with more parameters over ones with fewer parameters
- Therefore, an outside penalty accounting for our preference toward parsimony (e.g., in terms of number of components) needs to be incorporated in the model
- Hannig and Lee (2009) developed model selection in the GFI paradigm for wavelet regression and Lai et al. (2015) did it for ultra-high dimensional regression
- A fiducial factor is available, akin to a Bayes factor
- An outside penalty tailored towards mixture distributions could be derived, and, perhaps, some notion like a BIC difference (Raftery, 1995) can give us an indication of strength of a particular model

Determining the Number of Components

- A generalized fiducial model selection criterion could be used to determine the number of components, k
- We might include k in the parameter vector and find that generalized fiducial distribution
- Big challenge with this is that we are looking at deriving generalized fiducial quantities for parameters of varying dimensions
- A possibility is to use an extension of a Bernoulli factory (Łatuszyński et al., 2011), which uses martingale approaches to simulate a Bernoulli variable with success probability $f(p)$ from independent Bernoulli variables with success probability p
- Here, $p \in \mathcal{P} \subseteq [0, 1]$ is unknown, but $f : \mathcal{P} \to [0, 1]$ is known
- A Bernoulli factory could be used in an algorithm where $f(p)$ is the probability that a component is "born" or "dies", or we might consider developing something along the lines of a "multinoulli factory," where we simulate a multinoulli variable with success probability $f(\mathbf{p})$ from independent multinoulli variables with success probability $\mathbf{p}$

Computing

- Quick search of all `R` packages on `CRAN` yields only four packages with "fiducial" in the package name, each of which is focused on a specific class of models (e.g., logistic regression or normal linear mixed models), although other packages have some limited fiducial capabilities
- A realistic goal is to develop flexible, fiducial-based mixture functions for which we could employ `S3` methods
  - A pipe dream is to develop a comprehensive fiducial modeling architecture akin to `Stan`
- Generating observations from generalized fiducial distributions for conducting GFI is often computationally intensive, so efficiency in computational routines will be important

References I

Efron, B. (1998). R. A. Fisher in the 21st Century. *Statistical Science*, *13*(2), 95–114.

Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London, Series A*, *222*, 309–368.

Hannig, J. (2009). On Generalized Fiducial Inference. *Statistica Sinica*, *19*(2), 491–544.

Hannig, J., Iyer, H., Lai, R. C. S., & Lee, T. C. M. (2016). Generalized Fiducial Inference: A Review and New Results. *Journal of the American Statistical Association*, *111*(515), 1346–1361.

Hannig, J., & Lee, T. C. M. (2009). Generalized Fiducial Inference for Wavelet Regression. *Biometrika*, *96*(4), 847–860.

Izenman, A. J., & Sommer, C. J. (1988). Philatelic Mixtures and Multimodal Densities. *Journal of the American Statistical Association*, *83*(404), 941–953.

Lai, R. C. S., Hannig, J., & Lee, T. C. M. (2015). Generalized Fiducial Inference for Ultra-High Dimensional Regression. *Journal of the American Statistical Association*, *111*(510), 760–772.

Łatuszyński, K., Kosmidis, I., Papaspiliopoulos, O., & Roberts, G. O. (2011). Simulating Events of Unknown Probabilities via Reverse Time Martingales. *Random Structures & Algorithms*, *38*(4), 441–452.

Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, *25*, 111–163.

Contact Information

 derek.young@uky.edu

 http://young.as.uky.edu

 https://github.com/dsy109