SCIENCE CHINA
Mathematics

• ARTICLES •

https://doi.org/10.1007/s11425-021-2086-5

A fiducial approach to the nonparametric deconvolution problem: The discrete case

Yifan Cui[1,*] & Jan Hannig[2]

[1]*Center for Data Science, Zhejiang University, Hangzhou* 310058*, China;*
[2]*Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC* 27599*, USA*

Email: cuiyf@zju.edu.cn, jan.hannig@unc.edu

Received November 3, 2021; accepted January 11, 2023; published online August 14, 2024

Abstract Fiducial inference is applied to nonparametric g-modeling in the discrete case. We propose a computationally efficient algorithm to sample from the fiducial distribution and use the generated samples to construct point estimates and confidence intervals. We study the theoretical properties of the fiducial distribution and perform extensive simulations in various scenarios. The proposed approach gives rise to good statistical performance in terms of the mean squared error of point estimators and coverage of confidence intervals. Furthermore, we apply the proposed fiducial method to estimate the probability of each satellite site being malignant using gastric adenocarcinoma data with 844 patients.

Keywords confidence intervals, empirical Bayes, fiducial inference, nonparametric deconvolution

MSC(2020) 62G05, 62G09

Citation: Cui Y F, Hannig J. A fiducial approach to the nonparametric deconvolution problem: The discrete case. Sci China Math, 2024, 67, https://doi.org/10.1007/s11425-021-2086-5

1 Introduction

Efron [13, 14] and Narasimhan and Efron [38] studied the following important problem: an unknown distribution function $F(\theta)$ yields unobservable realizations $\Theta_1, \Theta_2, \ldots, \Theta_n$, and each Θ_i produces an observable value X_i according to a known probability mechanism. The goal is to estimate the unknown distribution function from the observed data. In this paper, we aim to provide a generalized fiducial solution to the same problem in the case where X_i given Θ_i follows a discrete distribution that has the known probability mass function g_i and the distribution function G_i. Following [14], we use the term "deconvolution" here, as the marginal distribution of X_i admits the following form; see also [14, (6)] and [38, (7)]:

$$\int g_i(x_i|\theta_i)dF_i(\theta_i). \tag{1.1}$$

Efron [14] proposed an empirical Bayes deconvolution approach to estimating the distribution of Θ from the observed sample $\{X_i, i = 1, \ldots, n\}$, where the only requirement is a known specification of

* Corresponding author

the distribution for X_i given Θ_i. The empirical Bayes deconvolution, since its development, has seen tremendous success in many scientific applications, including causal inference [29], single-cell analysis [53], cancer study [23,48], clinical trials [46,47], and many other fields [8]. Moreover, for classic Bayesian data analysis, as noted in [41, p. 19] and [22], a single distribution prior may sometimes be unsuitable and hence the prior choice is dubious. Efron's empirical Bayes deconvolution would be one of the alternatives since the obtained estimator of the distribution of Θ can be used as a prior distribution to produce posterior approximations [38]. The discrete deconvolution problem is common in practice, e.g., the famous missing species problem is an example of the Poisson deconvolution, and many studies in gene expression analysis use the Poisson or negative binomial distribution [40,45].

Fiducial inference can be traced back to Fisher [18, 19], who introduced the concept as a potential replacement for the Bayesian posterior distribution. Hannig et al. [25, 26] showed that fiducial distributions can be related to empirical Bayes methods, which are widely used in large-scale parallel inference problems [12,15,16]. Efron [11] pointed out that objective Bayes theories also have connections with fiducial inference. Other fiducial-related approaches include Dempster-Shafer theory [7, 9, 27, 36], inferential models [32–35], confidence distributions [28, 43, 44, 49, 55–57], and higher-order likelihood expansions and implied data-dependent priors [20, 21].

We propose a novel fiducial approach to modeling the distribution function F of Θ nonparametrically. In particular, we propose a computationally efficient algorithm to sample from the *generalized fiducial distribution* (GFD) (i.e., a distribution on a set of distribution functions) and use the generated samples to construct statistical procedures. The pointwise median of the GFD is used as a point estimate, and appropriate quantiles of the GFD evaluated at a given point provide pointwise confidence intervals. We also study the theoretical properties of the fiducial distribution. Extensive simulations in various scenarios show that the proposed fiducial approach is a good alternative to existing methods such as Efron's g-modeling. We apply the proposed fiducial approach to intestinal surgery data to estimate the probability of each satellite site being malignant for the patient; see [14] for the empirical Bayes approach. The resulting fiducial estimate of the distribution function reflects the observed patterns of raw data.

The rest of this paper is organized as follows. In Section 2, we present the mathematical framework for the fiducial approach to the nonparametric deconvolution problem. In Section 3, we establish an asymptotic theory, which verifies the frequentist validity of the proposed fiducial approach. Extensive simulation studies are presented in Section 4. We also illustrate our method using intestinal surgery data in Section 5. The paper concludes with a discussion of future work in Section 6. Some needed technical results are provided in the appendix.

2 Methodology

2.1 Data-generating equation

In this subsection, we first explain the definition of a GFD and then demonstrate how to apply it to the deconvolution problem. We start by expressing the relationship between the data X_i and the parameter Θ_i using

$$X_i = G_i^{-1}(U_i, \Theta_i), \quad \Theta_i = F^{-1}(W_i), \quad i = 1, \ldots, n, \tag{2.1}$$

where U_i and W_i are independent and identically distributed (i.i.d.) Unif$(0,1)$, $G_i(\cdot, \theta_i)$ are known distribution functions of discrete random variables supported on integers, G_i are defined and non-increasing in $\theta_i \in \mathcal{S} \subset \mathbb{R}$ for all i, and F is the unknown distribution function with the support in the interval $\mathcal{S}$. We are interested in estimating the unknown distribution function $F(\theta)$.

Recall that $F^{-1}(w) = \inf\{\theta : F(\theta) \geqslant w\}$ [3, p. 54], and $F^{-1}(w) = \theta$ if and only if $F(\theta) \geqslant w > F(\theta - \epsilon)$ for all $\epsilon > 0$. We define $G_i^*(x_i, u_i) = \sup\{\theta : G_i(x_i, \theta) \geqslant u_i\}$ with the usual understanding that $\sup \emptyset$ is smaller than all the elements of $\mathcal{S}$.

Remark 2.1. To facilitate a better understanding, we provide two examples for $G_i^*(x_i, u_i)$ here. If X follows a binomial distribution, G_i is the cumulative distribution function (CDF) of the binomial

distribution with the number of trials m_i, and $G_i^*(x_i, u_i)$ is the $(1-u_i)$ quantile of Beta(x_i+1, m_i-x_i); if X follows a Poisson distribution, G_i is the CDF of the Poisson distribution, and $G_i^*(x_i, u_i)$ is the $(1-u_i)$ quantile of Gamma$(x_i+1, 1)$.

If $G_i(\cdot, \theta_i)$ is continuous in θ_i, $G_i^*(x_i, u_i)$ is the solution (in θ_i) to the equation $G_i(x_i, \theta_i) = u_i$. Recall that the observed data points x_i are integers.

Lemma 2.2. *$x_i = G_i^{-1}(u_i, \theta_i)$ if and only if $G_i^*(x_i-1, u_i) < \theta_i \leqslant G_i^*(x_i, u_i)$.*

Combining $G_i^*(x_i-1, u_i) < \theta_i \leqslant G_i^*(x_i, u_i)$ and $F(\theta_i - \epsilon) < w_i \leqslant F(\theta_i)$ for all ϵ, consequently, we see that the inverse of the data-generating equation (2.1) is

$$Q_{\boldsymbol{x}}(\boldsymbol{u}, \boldsymbol{w}) = \{F : F(G_i^*(x_i-1, u_i)) < w_i \leqslant F(G_i^*(x_i, u_i)),\ i = 1, \ldots, n\}. \tag{2.2}$$

Remark 2.3. Note that given $\boldsymbol{x}$, $\boldsymbol{u}$ and $\boldsymbol{w}$, $Q_{\boldsymbol{x}}(\boldsymbol{u}, \boldsymbol{w})$ is a set of CDFs.

Lemma 2.4. *$Q_{\boldsymbol{x}}(\boldsymbol{u}, \boldsymbol{w}) \neq \emptyset$ if and only if $\boldsymbol{u}$ and $\boldsymbol{w}$ satisfy the condition that*

$$\textit{whenever } G_i^*(x_i, u_i) \leqslant G_j^*(x_j-1, u_j), \textit{ then } w_i < w_j. \tag{2.3}$$

A GFD is obtained by inverting the data-generating equation, and Hannig et al. [26] proposed a general definition of the GFD. However, in order to simplify the presentation, we use an earlier, less general version from [25]. These two definitions are equivalent for the models considered here. Suppose that $(\boldsymbol{U}^\star, \boldsymbol{W}^\star)$ are uniformly distributed on the set $\{(\boldsymbol{u}^\star, \boldsymbol{w}^\star) : Q_{\boldsymbol{x}}(\boldsymbol{u}^\star, \boldsymbol{w}^\star) \neq \emptyset\}$. A GFD is then the distribution of any element of the random set $\overline{Q_{\boldsymbol{x}}(\boldsymbol{U}^\star, \boldsymbol{W}^\star)}$, where the closure is in the weak topology on the space of probability measures on $\mathcal{S}$ and the element is selected so that it is measurable, i.e., a random distribution function on $\mathcal{S}$. Given the observed data $(x_1, \ldots, x_n)$, we define the random functions F^U and F^L as, for each $\theta \in \mathcal{S}$ and $(\boldsymbol{U}^\star, \boldsymbol{W}^\star)$,

$$F^U(\theta) \equiv \min\{W_i^\star \text{ for all } i \text{ such that } \theta < G_i^*(x_i-1, U_i^\star)\}$$

and

$$F^L(\theta) \equiv \max\{W_i^\star \text{ for all } i \text{ such that } \theta \geqslant G_i^*(x_i, U_i^\star)\},$$

where $\min\emptyset = 1$ and $\max\emptyset = 0$. These functions are clearly non-decreasing and right-continuous. Note that if $Q_{\boldsymbol{x}}(\boldsymbol{U}^\star, \boldsymbol{W}^\star) \neq \emptyset$, Portmanteau's theorem [1] and (2.2) imply that a distribution function $F \in \overline{Q_{\boldsymbol{x}}(\boldsymbol{U}^\star, \boldsymbol{W}^\star)}$ if and only if $F^L(\theta) \leqslant F(\theta) \leqslant F^U(\theta)$ for all $\theta \in \mathcal{S}$. Thus the functions F^U and F^L will be called the upper and lower fiducial bounds throughout. A sample from F^U and F^L can be used to perform estimation and inference for the unknown distribution function $F(\theta)$ in the same way that posterior samples are used in the Bayesian context. We generate the realizations of F^U and F^L by a novel Gibbs sampler in the next section.

2.2 Gibbs sampling and GFD-based inference

We need to generate $(\boldsymbol{U}^\star, \boldsymbol{W}^\star)$ from the standard uniform distribution on a set described by (2.3), which is achieved using a Gibbs sampler. For each fixed i, denote random vectors with the i-th observation removed by $(\boldsymbol{U}^\star_{[-i]}, \boldsymbol{W}^\star_{[-i]})$. If $(\boldsymbol{U}^\star, \boldsymbol{W}^\star)$ satisfy the constraint (2.3), so do $(\boldsymbol{U}^\star_{[-i]}, \boldsymbol{W}^\star_{[-i]})$. The proposed Gibbs sampler is based on the conditional distribution of

$$(U_i^\star, W_i^\star) \mid \boldsymbol{U}^\star_{[-i]}, \boldsymbol{W}^\star_{[-i]}, \tag{2.4}$$

which is a bivariate uniform distribution on a set A, a disjoint union of small rectangles. The beginnings and ends of the rectangles' bases are the neighboring points in the set $\bigcup_{j\neq i}\{G_i^*(x_i, U_j^\star), G_i^*(x_i-1, U_j^\star)\}$, while the corresponding location and height on the vertical axis are determined by (2.3). Details are described in Algorithm 1, and a visualization of the rectangles is shown in Figure 1. Each marginal conditional distribution is supported on the entire $\mathcal{S}$, and therefore we expect the proposed Gibbs sampler to mix well.

Algorithm 1 Pseudo algorithm for the fiducial Gibbs sampler

Input: Dataset, e.g., (m_i, x_i), $i = 1, \ldots, n$ for binomial data,
starting vectors $\boldsymbol{u}$ and $\boldsymbol{w}$ of length n,
n_{MCMC}, n_{burn}, and the vector $\boldsymbol{\theta}_{\text{grid}}$ of length n_{grid}.

for $i = 1$ **to** n **do**
  $\boldsymbol{\theta}_L[i] = G_i^*(x_i - 1, \boldsymbol{u}[i])$, $\boldsymbol{\theta}_U[i] = G_i^*(x_i, \boldsymbol{u}[i])$;
end
Run the Gibbs sampler using the initial values $\boldsymbol{u}$, $\boldsymbol{w}$, $\boldsymbol{\theta}_L$ and $\boldsymbol{\theta}_U$;
for $j = 1$ **to** $n_{\text{burn}} + n_{\text{MCMC}}$ **do**
  for $i = 1$ **to** n **do**
    $\boldsymbol{u}^0 = \boldsymbol{u}[-i]$, $\boldsymbol{w}^0 = \boldsymbol{w}[-i]$, $\boldsymbol{\theta}_L^0 = \boldsymbol{\theta}_L[-i]$, $\boldsymbol{\theta}_U^0 = \boldsymbol{\theta}_U[-i]$;
    $\boldsymbol{u}_L^{\text{pre}} = G_i(x_i, \boldsymbol{\theta}_L^0)$, $\boldsymbol{u}_U^{\text{pre}} = G_i(x_i - 1, \boldsymbol{\theta}_U^0)$;
    Sort $\boldsymbol{u}^{\text{pre}} = (\boldsymbol{u}_L^{\text{pre}}, \boldsymbol{u}_U^{\text{pre}}, 0, 1)$, denoted by $\boldsymbol{u}^{\text{sort}}$;
    Sort $(\boldsymbol{w}^0, 1(n-1), 1, 0)$ according to the order of $\boldsymbol{u}^{\text{pre}}$, denoted by $\boldsymbol{w}_U^\star$,
      where $1(n-1)$ is a vector with elements 1 of length $n-1$;
    Sort $(0(n-1), \boldsymbol{w}^0, 1, 0)$ according to the order of $\boldsymbol{u}^{\text{pre}}$, denoted by $\boldsymbol{w}_L^\star$,
      where $0(n-1)$ is a vector with elements 0 of length $n-1$;
    $\boldsymbol{w}_U^{\text{pre}} = \text{cummin}(\boldsymbol{w}_U^\star)$, $\boldsymbol{w}_L^{\text{pre}} = \text{rev}(\text{cummax}(\text{rev}(\boldsymbol{w}_L^\star)))$;
    Take the component-wise difference of $\boldsymbol{u}^{\text{sort}}$, denoted by $\boldsymbol{u}^{\text{diff}}$;
    for $k = 1$ **to** $2n - 1$ **do**
      $\boldsymbol{w}^{\text{diff}}[k] = \boldsymbol{w}_U^{\text{pre}}[k] - \boldsymbol{w}_L^{\text{pre}}[k+1]$;
    end
    Sample $i^\star \in \{1, \ldots, 2n-1\}$ with probability $\propto \boldsymbol{u}^{\text{diff}} \cdot \boldsymbol{w}^{\text{diff}}$;
    Sample a and b from independent Unif$(0, 1)$, and set
      $u = \boldsymbol{u}^{\text{sort}}[i^\star] + \boldsymbol{u}^{\text{diff}}[i^\star] \cdot a$, $w = \boldsymbol{w}_U^{\text{pre}}[i^\star] - \boldsymbol{w}^{\text{diff}}[i^\star] \cdot b$,
      $\theta_L = G_i^*(x_i - 1, u)$, $\theta_U = G_i^*(x_i, u)$,
      $\boldsymbol{u}[i] = u$, $\boldsymbol{w}[i] = w$, $\boldsymbol{\theta}_L[i] = \theta_L$, $\boldsymbol{\theta}_U[i] = \theta_U$;
  end
  Generate n i.i.d. Unif$(0, 1)$ and sort them according to the order of $\boldsymbol{w}$, denoted by $\boldsymbol{w}^*$. Replace $\boldsymbol{w}$ with $\boldsymbol{w}^*$, i.e., $\boldsymbol{w} = \boldsymbol{w}^*$;
end
Evaluate the upper and lower bounds on a grid of values $\boldsymbol{\theta}_{\text{grid}}$ for each Markov chain Monte Carlo (MCMC) sample after burn-in, indexed by l;
for $j = 1$ **to** n_{grid} **do**
  $F_l^L(\boldsymbol{\theta}_{\text{grid}}[j]) = \max(\boldsymbol{w}[\boldsymbol{\theta}_U \leqslant \boldsymbol{\theta}_{\text{grid}}[j]])$;
  $F_l^U(\boldsymbol{\theta}_{\text{grid}}[j]) = \min(\boldsymbol{w}[\boldsymbol{\theta}_L \geqslant \boldsymbol{\theta}_{\text{grid}}[j]])$;
end
return The fiducial samples F_l^U and F_l^L evaluated on $\boldsymbol{\theta}_{\text{grid}}$.

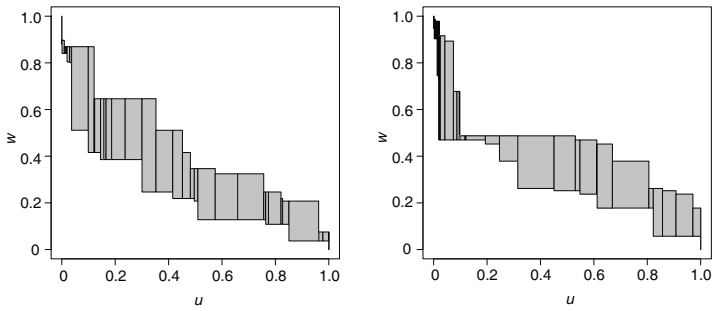

Figure 1 A visualization of rectangles in Algorithm 1 for the final step of the Gibbs sampler ($j = n_{\text{burn}} + n_{\text{MCMC}}, i = n$) for the toy example described in Subsection 2.3. The last (U_n, W_n) was generated from the gray area, with the rectangles indicated by black lines. Each panel represents a different starting value

The proposed Gibbs sampler needs starting points, and we consider two potential initializations. The first starts with randomly generated $(\boldsymbol{U}^\star, \boldsymbol{W}^\star)$ from independent Unif$(0, 1)$ and reorders $\boldsymbol{W}^\star$ so that the constraint (2.3) is satisfied. The second starts with a deterministic Θ, e.g., $p = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n m_i}$ for binomial data $X_i \sim \text{Bin}(m_i, p)$ and $\lambda = \frac{\sum_{i=1}^n x_i}{n}$ for Poisson data $X_i \sim \text{Poi}(\lambda)$. As these two starting points are very different, they can be used to monitor convergence. To streamline our presentation, in Sections 4 and 5,

we present numerical results using the first initialization.

Algorithm 1 outputs two distribution functions necessary for constructing the proposed mixture and conservative confidence intervals. In the rest of this paper, we denote Monte Carlo realizations of the lower and upper fiducial bounds by F_l^L and F_l^U, respectively, where $l = 1, \ldots, n_{\text{MCMC}}$, with n_{MCMC} being the number of fiducial samples.

We propose to use the median of the $2n_{\text{MCMC}}$ samples $\{F_l^L(\theta), F_l^U(\theta), l = 1, \ldots, n_{\text{MCMC}}\}$ as a point estimator of the distribution function $F(\theta)$. We construct two types of pointwise confidence intervals, conservative and mixture, using appropriate quantiles of fiducial samples [25]. In particular, the 95% conservative confidence interval is formed by taking the empirical 0.025 quantile of $\{F_l^L(\theta), l = 1, \ldots, n_{\text{MCMC}}\}$ as the lower limit and the empirical 0.975 quantile of $\{F_l^U(\theta), l = 1, \ldots, n_{\text{MCMC}}\}$ as the upper limit. The lower and upper limits of the 95% mixture confidence interval are formed by taking the empirical 0.025 and 0.975 quantiles of $\{F_l^L(\theta), F_l^U(\theta), l = 1, \ldots, n_{\text{MCMC}}\}$, respectively.

2.3 Further illustration with a simulated dataset

To streamline our presentation, we take the binomial case as our running example in the following, i.e., the observed data are (m_i, x_i), and $X_i \sim \text{Bin}(m_i, P_i)$, $i = 1, \ldots, n$, where $P_i \in \mathcal{S} = [0, 1]$ plays the role of Θ_i. We also provide the details of the proposed approach and some examples of the Poisson data in the appendix.

We present a toy example to demonstrate the proposed fiducial approach. Suppose that $F \sim \text{Beta}(5, 5)$. The number of trials is $m_i = 20$, $i = 1, \ldots, n$. The sample size of the simulated binomial data is $n = 20$. The fiducial estimates are based on 10,000 iterations after 1,000 burn-in iterations.

(a) and (b) in Figure 2 present the last MCMC sample of the lower fiducial bound $F_l^L(p)$ (in the blue line) and upper fiducial bound $F_l^U(p)$ (in the red line) for the two starting points, respectively. As the fiducial distribution reflects the uncertainty, we do not expect every single fiducial curve to be close to the true CDF (in the black line). Furthermore, (a) and (b) in Figure 3 present the mixture confidence interval (with the blue line for the lower limit and the red line for the upper limit) and conservative confidence interval (with the cyan line for the lower limit and the magenta line for the upper limit), computed from the MCMC sample for the two starting points, respectively. In addition, we plot the point estimates of the proposed approach along with Efron's g-modeling. The brown curve is the fiducial point estimate $\widehat{F}(p)$. The dashed curve is the point estimate of $F(p)$ for Efron's g-modeling without bias correction. Efron's confidence intervals with bias correction look almost the same as without correction, and thus are omitted in the figures. As can be seen, the proposed fiducial point estimator and confidence intervals capture the shape of the true CDF pretty well.

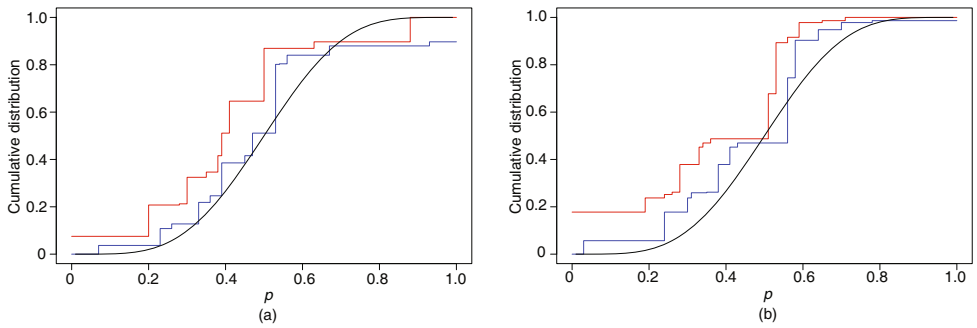

Figure 2 (Color online) The last MCMC sample from the GFD, with each panel corresponding to a different starting point. The blue curve is a realization of the lower fiducial bound $F^L(p)$, the red curve is a realization of the upper fiducial bound $F^U(p)$, and the black curve is the true $F(p)$

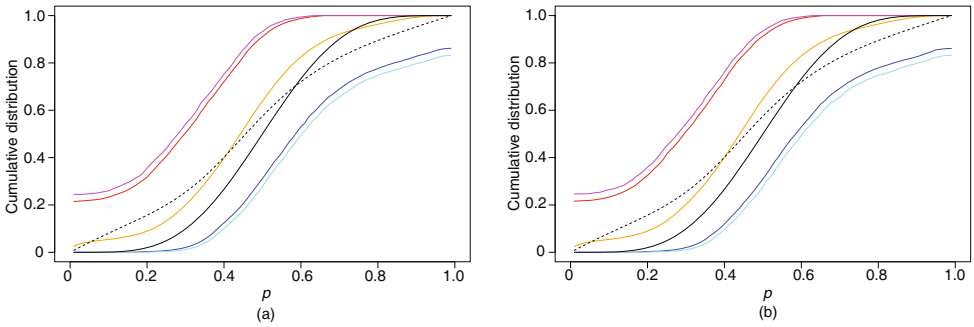

Figure 3 (Color online) Point estimates and 95% confidence intervals (CIs) for $F(p)$ given a fixed simulated dataset, with each panel corresponding to a different starting point of the MCMC chain. The orange curve is the fiducial point estimate $\widehat{F}(p)$. The dashed curve is the point estimate of $F(p)$ for Efron's g-modeling. The black curve is the true $F(p)$. The blue and red curves are the lower and upper limits of mixture CIs, respectively. The cyan and magenta curves are the lower and upper limits of conservative CIs, respectively

3 Theoretical results

Recall that the GFD is a data-dependent distribution, which is defined for every fixed dataset $\boldsymbol{x}$. It can be made into a random measure in the same way as one defines the usual conditional distribution, by plugging random variables $\boldsymbol{X}$ into the observed dataset. In this section, we study the asymptotic behavior of this random measure for $X_i \sim \mathrm{Bin}(m_i, P_i)$ when the rate of m_i is much faster than n, i.e., the following assumption holds.

Assumption 3.1. $\lim_{n\to\infty} n^4(\log n)^{1+\epsilon}/(\min_{i=1,\ldots,n} m_i) = 0$ *for any* $\epsilon > 0$.

We provide a central limit theorem for $F^L(p)$. The proof of Theorem 3.2 is deferred to the appendix. A similar result holds for $F^U(p)$.

Theorem 3.2. *Suppose that the true CDF* F *is absolutely continuous with a bounded density. Based on Assumption* 3.1*,*

$$n^{1/2}\{F^L(\cdot) - \widehat{F}_n(\cdot)\} \to B_F(\cdot) \tag{3.1}$$

in distribution on Skorokhod space $\mathcal{D}[0,1]$ *in probability, where*

$$\widehat{F}_n(s) \equiv \frac{1}{n}\sum_{i=1}^{n} I[P_i \leqslant s] \tag{3.2}$$

is the oracle empirical CDF constructed from the unobserved P_i*, which were used to generate the observed* X_i*,* $i = 1, \ldots, n$*, and* $B_F(\cdot)$ *is a mean-zero Gaussian process with covariance*

$$\mathrm{cov}(B_F(s), B_F(t)) = F(t \wedge s) - F(t)F(s).$$

Notice that the stochastic process on the left-hand side of (3.1) is naturally in $\mathcal{D}[0,1]$, the space of functions on $[0,1]$ that are right-continuous and have left limits. Distances on $\mathcal{D}[0,1]$ are measured using Skorokhod's metric, which makes it into a Polish space [1]. To understand the mode of convergence used here, note that there are two sources of randomness present. One is from the fiducial distribution itself, which is derived from each fixed data set. The other is the usual randomness of the data. Thus, (3.1) can be interpreted as

$$\rho(n^{1/2}\{F^L(\cdot) - \widehat{F}_n(\cdot)\}, B_F(\cdot)) \xrightarrow{\mathrm{pr}} 0, \tag{3.3}$$

where ρ is any metric metrizing weak convergence of probability measures on the Polish space $\mathcal{D}[0,1]$, e.g., the Lévy-Prokhorov or Dudley metric [50]. The distribution of $n^{1/2}\{F^L(\cdot) - \widehat{F}_n(\cdot)\}$ in the argument of ρ is the fiducial distribution, induced by the randomness of $(\boldsymbol{U}^\star, \boldsymbol{W}^\star)$ with the data X_i and P_i being fixed.

Consequently, the left-hand side of (3.3) is a function of X_i and P_i, and the convergence in probability is based on the distribution of the data.

Theorem 3.2 establishes a Bernstein-von Mises theorem for the fiducial distribution under Assumption 3.1, which implies that the confidence intervals described in Section 2 have asymptotically correct coverage. Moreover, Theorem 3.2 provides a sufficient condition for $n^{1/2}$-estimability of the binomial probability parameter's distribution function. While Assumption 3.1 is pretty stringent, it does not seem likely that one can establish a unified asymptotic property of the proposed fiducial approach under a general scheme. This is best seen by the fact that in the binomial case, if m_i's are uniformly bounded, there is not enough information in the data to consistently estimate the underlying distribution function of P.

Interestingly, looking at the fiducial solution reveals an interesting connection to a different statistical problem [6]. Recall that the quantities $P_i,\ i = 1, \ldots, n$ are only known to be inside random intervals $(G_i^*(x_i - 1, U_i^\star), G_i^*(x_i, U_i^\star)]$. Therefore, the statistical problem we study here bears similarities to nonparametric estimation under Turnbull's general censoring scheme [10,52], in which case there is no unified theory of the nonparametric maximum likelihood estimator, but properties are investigated under various special and challenging cases such as $n^{1/2}$-convergence for right-censored data [2], and $n^{1/3}$-convergence for current status data [24].

Remark 3.3. Suppose $m_i \equiv m$. Notice that

$$\mathrm{pr}(F \in Q_x(U^*, W^*)) \propto \prod_{i=1}^{n} \left[\int_0^1 \binom{m}{x_i} p^{x_i}(1-p)^{m-x_i} dF(p) \right]$$

is proportional to the nonparametric likelihood function. A simple calculation shows that maximizing the scaled fiducial probability in its limit provides exactly the underlying true CDF. Detailed derivations and further discussions of this observation are provided in the appendix.

4 Numerical experiments

We perform simulation studies to compare the frequentist properties of the proposed fiducial confidence intervals with Efron's g-modeling [14, 17], the nonparametric bootstrap, and a nonparametric Bayesian approach [42]. In each scenario, we first generate p_i, $i = 1, \ldots, n$ from the distribution function F. Then we draw X_i from the binomial distribution $\mathrm{Bin}(m_i, p_i)$, where m_i is described in Subsection 4.1. The simulations are replicated 500 times in each scenario.

For the proposed method as well as other existing methods, we choose the grid $p = 0.01, 0.02, \ldots, 0.99$ following [38]. The fiducial estimates are based on 2,000 iterations of the Gibbs sampler after 500 burn-in iterations. Efron's g-modeling is implemented using the R package `deconvolveR` [17]. We use the default value of 5 for the degree of the splines. We consider the regularization strategy with the default value $c_0 = 1$. For the nonparametric bootstrap, we first obtain the maximum likelihood estimates $\hat{p}_i = x_i/m_i$. We then construct the empirical CDF as the point estimator and used $B = 1{,}000$ bootstrap samples of $\hat{p}_i$ to construct confidence intervals. We also consider a fully Bayesian approach, i.e., the Dirichlet process mixture of Beta-binomial distributions, which gives more flexibility than a Beta-binomial model [41, p. 19]. Default values for the prior parameters in the R package `dirichletprocess` [42] are used. The Bayesian estimates are based on 2,000 MCMC samples after 500 burn-in iterations.

4.1 Simulation settings

We start with the following two scenarios with m_i being the same across individuals.

Scenario 1. We consider the same setting as Subsection 2.3. Let $\Theta \sim \mathrm{Beta}(5, 5)$, and the number of trials is $m_i = 20$, $i = 1, \ldots, n$. The sample size n of the simulated binomial data is set to 50.

Scenario 2. Let $\Theta \sim 0.5\mathrm{Beta}(10, 30) + 0.5\mathrm{Beta}(30, 10)$, and the number of trials is $m_i = 20$, $i = 1, \ldots, n$. The sample size n of the simulated binomial data is set to 50.

Next, we consider three complex settings from [58].

Scenario 3 (Beta density)**.** Let $\Theta \sim \text{Beta}(8, 8)$, and m_i's are integers sampled uniformly between 100 and 200. The sample size n of the simulated binomial data equals 100.

Scenario 4 (A multimodal distribution)**.** Let $\Theta \sim 0.5\text{Beta}(60, 10) + 0.5\text{Beta}(10, 60)$, and the number of trials is $m_i = 100$, $i = 1, \ldots, n$. The sample size n of the simulated binomial data equals 100.

Scenario 5 (Truncated exponential)**.** Let $\Theta \sim \text{Exp}(8)$ truncated at 1, and the number of trials is $m_i = 100$, $i = 1, \ldots, n$. The sample size n of the simulated binomial data equals 200.

4.2 Numerical results

In this subsection, we first compare the mean squared errors (MSEs) of different methods for estimating $F(p)$ in the five scenarios. The numerical results for

$$p = 0.15, 0.25, 0.50, 0.75, 0.85$$

in each scenario are presented in Table 1. We see that the MSEs of the proposed fiducial point estimates are as good as, and sometimes smaller than, those of competing methods.

Next, we present the coverage and average length of confidence intervals for various methods in Tables 2 and 3. Table 2 summarizes the coverage of 95% confidence intervals of various methods, and Table 3

Table 1 MSE ($\times 10^{-4}$) of point estimates for $F(p)$ in each scenario

Scenario	p	F	g	bc	BP	BA
1	0.15	5	40	39	17	1
	0.25	20	49	48	69	10
	0.50	51	47	48	71	78
	0.75	18	22	21	16	10
	0.85	5	13	12	4	1
2	0.15	41	149	149	145	9
	0.25	39	17	18	66	101
	0.50	49	32	33	54	53
	0.75	37	18	19	50	88
	0.95	46	86	84	33	12
3	0.15	0.14	8.45	8.16	0.15	0.04
	0.25	2.57	14.89	14.41	2.87	1.39
	0.50	22.85	26.94	27.04	24.68	21.92
	0.75	2.54	5.03	4.81	2.80	1.34
	0.85	0.16	1.46	1.38	0.15	0.05
4	0.15	22	49	49	23	27
	0.25	27	34	33	26	25
	0.50	25	24	25	26	26
	0.75	26	36	35	27	25
	0.85	20	52	52	26	25
5	0.15	11	10	10	11	10*
	0.25	6	10	10	6	6*
	0.50	1	5	4	1	1*
	0.75	0.13	1.84	1.66	0.12	0.10*
	0.85	0.05	0.76	0.68	0.05	0.04*

* The Bayesian results in Scenario 5 are reported based on 489 replications as 11 runs fail due to an error in the R package `dirichletprocess`.

summarizes the average length of these confidence intervals. "F" denotes the fiducial point estimates; "M" denotes the mixture GFD confidence intervals; "C" denotes the conservative GFD confidence intervals; "g" denotes Efron's g-modeling without bias correction; "bc" denotes Efron's g-modeling with bias correction; "BP" denotes the bootstrap method; "BA" denotes the Bayesian method.

For point estimators, overall, the proposed fiducial method and the Bayesian method outperform other methods. The performances of the proposed estimator and the Bayesian estimator are comparable, with the former outperforming the latter for medium values of p in Scenarios 1 and 2 and the latter outperforming the former for small and large values of p in Scenarios 1–3. For uncertainty quantification, we see that the GFD confidence intervals maintain or exceed the nominal coverage everywhere, while other methods often have coverage problems. In particular, Efron's confidence intervals and nonparametric bootstrap have substantial coverage problems close to the boundary, while the Bayesian method consistently underestimates the uncertainty, resulting in credible intervals that are too narrow.

It is not surprising that the GFD confidence intervals are often longer than other methods as the GFD approach aims to provide a conservative way to quantify uncertainty. As expected, the mean length of mixture GFD confidence intervals is slightly shorter than conservative GFD confidence intervals. A potential reason for the fiducial approach outperforming Efron's g-modeling in terms of coverage is that Efron's g-modeling relies on an exponential family parametric model and the proposed fiducial approach is nonparametric. Therefore, the proposed method is demonstrably robust to certain model mis-specifications, e.g., when the true model does not belong to an exponential family.

Table 2 Coverage (in percent) of 95% CIs for $F(p)$ in each scenario

Scenario	p	M	C	g	bc	BP	BA
1	0.15	99	100	25	27	74	88
	0.25	99	100	80	82	65	83
	0.50	100	100	94	94	91	84
	0.75	100	100	96	96	92	83
	0.85	100	100	90	91	54	89
2	0.15	98	99	1	1	27	97
	0.25	100	100	98	98	90	86
	0.50	98	98	96	96	97	45
	0.75	100	100	96	96	85	87
	0.85	97	99	31	31	79	97
3	0.15	100	100	4	5	13	28
	0.25	99	99	69	71	86	28
	0.50	99	99	89	89	96	27
	0.75	98	99	97	97	87	32
	0.85	99	100	97	98	12	31
4	0.15	99	100	48	49	95	59
	0.25	95	97	89	90	93	12
	0.50	95	96	93	93	95	6
	0.75	95	95	89	89	93	9
	0.85	99	99	75	75	91	59
5	0.15	99	99	95	95	93	18*
	0.25	97	98	95	95	93	21*
	0.50	98	98	98	98	89	20*
	0.75	98	99	100	100	36	20*
	0.85	99	100	100	100	17	11*

* The Bayesian results in Scenario 5 are reported based on 489 replications as 11 runs fail due to an error in the R package `dirichletprocess`

Table 3 Mean length ($\times 10^{-3}$) of 95% CIs for $F(p)$ in each scenario

Scenario	p	M	C	g	bc	BP	BA
1	0.15	123	139	101	101	84	28
	0.25	220	241	171	171	172	95
	0.50	426	465	246	246	272	279
	0.75	217	239	154	154	128	89
	0.85	121	137	82	81	42	28
2	0.15	243	266	66	66	185	101
	0.25	357	390	162	162	251	311
	0.50	305	330	206	206	275	90
	0.75	352	385	166	166	226	301
	0.85	245	268	143	143	133	107
3	0.15	35	42	46	46	4	2
	0.25	77	84	81	81	51	10
	0.50	243	259	168	168	195	33
	0.75	78	86	65	65	51	10
	0.85	35	42	27	26	4	2
4	0.15	240	259	125	125	179	101
	0.25	201	213	191	191	195	16
	0.50	193	202	188	188	196	0.02
	0.75	201	213	198	198	195	16
	0.85	240	259	178	178	173	98
5	0.15	160	171	125	125	126	18*
	0.25	115	124	114	114	94	13*
	0.50	45	50	74	73	35	6*
	0.75	20	24	37	36	7	1*
	0.85	17	20	23	22	3	1*

* The Bayesian results in Scenario 5 are reported based on 489 replications as 11 runs fail due to an error in the R package `dirichletprocess`

5 Intestinal surgery data

In this section, we consider an intestinal surgery study on gastric adenocarcinoma involving $n = 844$ cancer patients [23]. Resection of the primary tumor with appropriate dissection of surrounding lymph nodes is the foundation of curative care. In addition to the primary tumor, surgeons also remove satellite nodes for later testing. Efron's deconvolution is used to estimate the prior distribution of the probability of one satellite being malignant in this study [23].

The dataset consists of pairs (m_i, X_i), $i = 1, \ldots, n$, where m_i's are the numbers of satellites removed, and X_i's are the numbers of these satellites found to be malignant. m_i's vary from 1 to 69. Among all the cases, 322 have $X_i = 0$. For the rest of them, X_i/m_i has an approximate Unif$(0, 1)$ distribution [14]. We are interested in estimating the distribution function of the probability of one satellite being malignant. Following the model proposed in [14], we assume a binomial model, i.e., $X_i \sim \text{Bin}(m_i, P_i)$, where P_i is the i-th patient's probability of any satellite being malignant.

We compare the proposed mixture and conservative GFD confidence intervals to Efron's with and without bias correction, the bootstrap method, and the fully Bayesian approach. For all the methods, we use the grid $[0.01, 0.02, \ldots, 0.99]$ for the discretization of p. The fiducial and Bayesian estimates are based on 10,000 iterations after 1,000 burn-in iterations. Other tuning parameters for each method are chosen in the same way as Section 4.

The overall shapes of bootstrap and Bayesian confidence intervals are similar to the fiducial ones but much narrower. We provide bootstrap and Bayesian point estimates and 95% confidence intervals

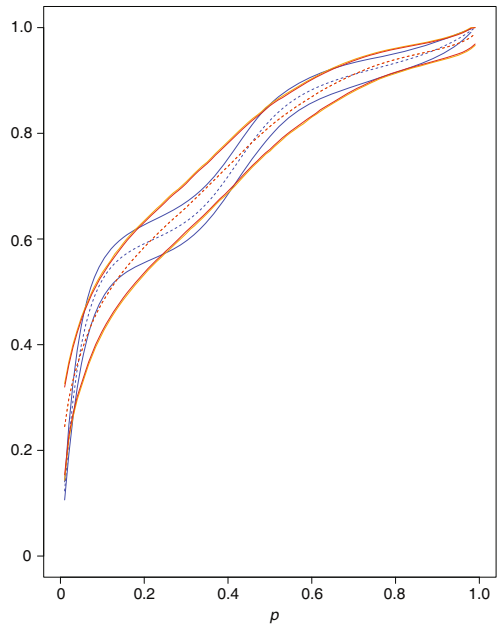

Figure 4 (Color online) Estimated CDF (dashed line) and 95% CIs for $F(p)$ of the GFD and Efron's g-modeling. The red and orange curves are mixture and conservative confidence intervals, respectively. The blue curve is Efron's confidence intervals without bias correction. Efron's confidence intervals with bias correction look almost the same as without correction, and thus are omitted

in the appendix. Figure 4 shows the point estimates and 95% confidence intervals of the distribution function F for the proposed GFD approach and Efron's g-modeling. Overall, the GFD confidence interval is more conservative. The GFD confidence intervals cover Efron's almost everywhere.

For the proposed fiducial approach, there is a large mode for the upper fiducial confidence interval near $p = 0$, which coincides with the fact that about 38% of the X_i's are 0 in the surgery data. However, the Bayesian method and Efron's g-modeling seem to quantify the uncertainty of this proportion to be lower. One exception is the nonparametric bootstrap, which estimates the point mass at zero to be 0.38 with a 95% confidence interval of $(0.35, 0.41)$. We note that this might be overestimated as $X_i = 0$ may correspond to a non-zero probability p, especially when m_i is small.

Moreover, the generalized fiducial confidence intervals provide us a unimodal density, while Efron's gives a bimodal density. We believe that the fiducial, bootstrap, and nonparametric Bayesian answers are more in line with Efron's observation that for those $X_i \neq 0$ in the surgery data, X_i/m_i has an approximate $\text{Unif}(0, 1)$ distribution [14].

6 Discussion

In this paper, we propose a prior-free approach to a nonparametric deconvolution problem and obtain valid point estimates and confidence intervals. This is accomplished through a novel algorithm to sample from the GFD. The median of the GFD is used as the point estimate, and appropriate quantiles of the GFD evaluated at a given p provide pointwise confidence intervals. We also study the theoretical properties of the fiducial distribution. Extensive simulations show that the proposed fiducial approach is a good alternative to existing methods such as Efron's g-modeling. We apply the proposed fiducial approach to intestinal surgery data to estimate the probability of each satellite site being malignant for patients.

We conclude by listing some open research problems as follows:

1. The proposed fiducial method seems to be a powerful nonparametric approach. It would be

interesting to implement it inside other statistical procedures, such as tree or random forest models, to include covariates [54].

2. This paper focuses on discrete data. The proposed approach can be extended to continuous data, such as Normal$(\Theta, 1)$, where Θ follows a distribution function F. This part is currently under investigation.

3. As we can see in simulations, the GFD approach is sometimes over-conservative. It could be possible to consider a different choice of fiducial samples, such as log-interpolation [4] or monotonic spline interpolation [5, 51].

4. The asymptotic distribution results in Section 3 only hold under Assumption 3.1. Investigating non-$n^{1/2}$-convergence should be a fruitful avenue of future research.

5. It should be possible to use the GFD in conjunction with various functional norms to construct simultaneous confidence bands [4, 31, 37].

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. U23A2064), Singapore Ministry of Education, U.S. National Institute of Health, and U.S. National Science Foundation. The authors are thankful to Professor Hari Iyer (U.S. National Institute of Standards and Technology) for helpful discussions. The authors are also thankful to the referees for their helpful comments, which led to an improved manuscript.

References

1 Billingsley P. Convergence of Probability Measures, 2nd ed. New York: Wiley-Interscience, 1999
2 Breslow N, Crowley J. A large sample study of the life table and product limit estimates under random censorship. Ann Statist, 1974, 2: 437–453
3 Casella G, Berger R L. Statistical Inference, 2nd ed. Pacific Grove: Duxbury Press, 2002
4 Cui Y, Hannig J. Nonparametric generalized fiducial inference for survival functions under censoring (with discussions and rejoinder). Biometrika, 2019, 106: 501–518
5 Cui Y, Hannig J. Rejoinder: 'Nonparametric generalized fiducial inference for survival functions under censoring'. Biometrika, 2019, 106: 527–531
6 Cui Y, Hannig J, Kosorok M. A unified nonparametric fiducial approach to interval-censored data. J Amer Statist Assoc, 2024, in press
7 Dempster A P. The Dempster-Shafer calculus for statisticians. Internat J Approx Reason, 2008, 48: 365–377
8 Dulek B. Empirical Bayes deconvolution based modulation discovery under additive noise. IEEE Trans Veh Technol, 2018, 67: 6668–6672
9 Edlefsen P T, Liu C, Dempster A P. Estimating limits from Poisson counting data using Dempster-Shafer analysis. Ann Appl Stat, 2009, 3: 764–790
10 Efron B. The two sample problem with censored data. In: Le Cam L M, Neyman J, eds. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Biology and Problems of Health. Berkeley: University of California Press, 1967, 831–853
11 Efron B. R. A. Fisher in the 21st century. Statist Sci, 1998, 13: 95–122
12 Efron B. Large-Scale Inference: Empirical Bayes Methods For Estimation, Testing, and Prediction. Cambridge: Cambridge Univ Press, 2012
13 Efron B. Two modeling strategies for empirical Bayes estimation. Statist Sci, 2014, 29: 285–301
14 Efron B. Empirical Bayes deconvolution estimates. Biometrika, 2016, 103: 1–20
15 Efron B. Bayes, oracle Bayes and empirical Bayes. Statist Sci, 2019, 34: 177–201
16 Efron B. Rejoinder: Bayes, oracle Bayes, and empirical Bayes. Statist Sci, 2019, 34: 234–235
17 Efron B, Narasimhan B. deconvolveR: Empirical Bayes estimation strategies. R package version 1.2-1. DOI:10.32614/CRAN.package.deconvolveR, 2020
18 Fisher R A. Inverse probability. Math Proc Cambridge Philos Soc, 1930, 26: 528–535
19 Fisher R A. The concepts of inverse probability and fiducial probability referring to unknown parameters. Proc Roy Soc Lond Ser A, 1933, 139: 343–348
20 Fraser D A S. Ancillaries and conditional inference. Statist Sci, 2004, 19: 333–369
21 Fraser D A S. Is Bayes posterior just quick and dirty confidence? Statist Sci, 2011, 26: 299–316
22 Gelman A, Carlin J B, Stern H S, et al. Bayesian Data Analysis. Boca Raton: CRC Press, 2013
23 Gholami S, Janson L, Worhunsky D J, et al. Number of lymph nodes removed and survival after gastric cancer resection: An analysis from the US Gastric Cancer Collaborative. J Amer Coll Surgeons, 2015, 221: 291–299

24 Groeneboom P, Wellner J A. Information Bounds and Nonparametric Maximum Likelihood Estimation. New York: Springer, 1992

25 Hannig J. On generalized fiducial inference. Statist Sinica, 2009, 19: 491–544

26 Hannig J, Iyer H, Lai R C S, et al. Generalized fiducial inference: A review and new results. J Amer Statist Assoc, 2016, 111: 1346–1361

27 Hannig J, Xie M. A note on Dempster-Shafer recombination of confidence distributions. Electron J Stat, 2012, 6: 1943–1966

28 Hjort N L, Schweder T. Confidence distributions and related themes. J Statist Plann Inference, 2018, 195: 1–13

29 Lee K, Small D S. Estimating the Malaria attributable fever fraction accounting for parasites being killed by fever and measurement error. J Amer Statist Assoc, 2019, 114: 79–92

30 Marchal O, Arbel J. On the sub-Gaussianity of the Beta and Dirichlet distributions. Electron Commun Probab, 2017, 22: 1–14

31 Martin R. Discussion of 'Nonparametric generalized fiducial inference for survival functions under censoring'. Biometrika, 2019, 106: 519–522

32 Martin R, Liu C. Inferential models: A framework for prior-free posterior probabilistic inference. J Amer Statist Assoc, 2013, 108: 301–313

33 Martin R, Liu C. Marginal inferential models: Prior-free probabilistic inference on interest parameters. J Amer Statist Assoc, 2015, 110: 1621–1631

34 Martin R, Liu C. Conditional inferential models: Combining information for prior-free probabilistic inference. J R Stat Soc Ser B Stat Methodol, 2015, 77: 195–217

35 Martin R, Liu C. Inferential Models: Reasoning with Uncertainty. Monographs on Statistics and Applied Probability. Boca Raton: CRC Press, 2015

36 Martin R, Zhang J, Liu C. Dempster-Shafer theory and statistical inference with weak beliefs. Statist Sci, 2010, 25: 72–87

37 Nair V N. Confidence bands for survival functions with censored data: A comparative study. Technometrics, 1984, 26: 265–275

38 Narasimhan B, Efron B. deconvolveR: A *G*-modeling program for deconvolution and empirical Bayes estimation. J Statist Softw, 2020, 94: 1–20, https://doi.org/10.18637/jss.v094.i11

39 Praestgaard J, Wellner J A. Exchangeably weighted bootstraps of the general empirical process. Ann Probab, 1993, 21: 2053–2086

40 Robinson M D, McCarthy D J, Smyth G K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 2010, 26: 139–140

41 Ross G J, Markwick D. dirichletprocess: An R package for fitting complex Bayesian nonparametric models. Https://cran.r-project.org/web/packages/dirichletprocess/vignettes/dirichletprocess.pdf, 2018

42 Ross G J, Markwick D. dirichletprocess: Build Dirichlet process objects for Bayesian modelling. R package version 0.4.2. DOI:10.32614/CRAN.package.dirichletprocess, 2023

43 Schweder T, Hjort N L. Confidence and likelihood. Scand J Stat, 2002, 29: 309–332

44 Schweder T, Hjort N L. Confidence, Likelihood, Probability. Cambridge: Cambridge Univ Press, 2016

45 Schwender H. siggenes: Multiple testing using SAM and Efron's empirical Bayes approaches. R package version 1.78.0. DOI:10.18129/B9.bioc.siggenes, 2024

46 Shen C, Li X. Using previous trial results to inform hypothesis testing of new interventions. J Biopharm Stat, 2018, 28: 884–892

47 Shen C, Li X. Towards more flexible false positive control in phase III randomized clinical trials. arXiv:1902.08229, 2019

48 Shen C, Xu H. Randomized phase III oncology trials: A survey and empirical Bayes inference. J Stat Theory Pract, 2019, 13: 49

49 Shen J, Liu R Y, Xie M. iFusion: Individualized fusion learning. J Amer Statist Assoc, 2020, 115: 1251–1267

50 Shorack G R. Probability for Statisticians. Heidelberg: Springer, 2017

51 Taraldsen G, Lindqvist B H. Discussion of 'Nonparametric generalized fiducial inference for survival functions under censoring'. Biometrika, 2019, 106: 523–526

52 Turnbull B W. The empirical distribution function with arbitrarily grouped, censored and truncated data. J R Stat Soc Ser B Stat Methodol, 1976, 38: 290–295

53 Wang J, Huang M, Torre E, et al. Gene expression distribution deconvolution in single-cell RNA sequencing. Proc Natl Acad Sci USA, 2018, 115: E6437–E6446

54 Wu S, Hannig J, Lee T. Uncertainty quantification in ensembles of honest regression trees using generalized fiducial inference. arXiv:1911.06177, 2019

55 Xie M, Liu R Y, Damaraju C V, et al. Incorporating external information in analyses of clinical trials with binary outcomes. Ann Appl Stat, 2013, 7: 342–368

56 Xie M, Singh K. Confidence distribution, the frequentist distribution estimator of a parameter: A review. Int Stat Rev, 2013, 81: 3–39
57 Xie M, Singh K, Strawderman W E. Confidence distributions and a unifying framework for meta-analysis. J Amer Statist Assoc, 2011, 106: 320–333
58 Zhang T, Liu J S. Nonparametric hierarchical Bayes analysis of binomial data via Bernstein polynomial priors. Canad J Statist, 2012, 40: 328–344

Appendix A Proofs

Proof of Lemma 2.2. Recall the definition of $G_i^{-1}(u_i,\theta_i) = \inf\{x_i : G_i(x_i,\theta_i) \geqslant u_i\}$. So $x_i = G_i^{-1}(u_i,\theta_i)$ if and only if $G_i(x_i,\theta_i) \geqslant u_i > G_i(x_i-\epsilon,\theta_i)$ for all $\epsilon > 0$. Then by the definition of G_i^*, it is further equivalent to $G_i^*(x_i-1,u_i) < \theta_i \leqslant G_i^*(x_i,u_i)$. □

Proof of Lemma 2.4. Sufficiency. If $Q_{\boldsymbol{x}}(\boldsymbol{u},\boldsymbol{w}) \neq \emptyset$ and $G_i^*(x_i,u_i) \leqslant G_j^*(x_j-1,u_j)$, then we know that

$$w_i \leqslant F(G_i^*(x_i,u_i)) \leqslant F(G_j^*(x_j-1,u_j)) < w_j.$$

Necessity. We prove it by contradiction. If $Q_{\boldsymbol{x}}(\boldsymbol{u},\boldsymbol{w})$ is empty, then there must exist indices i and j such that $(G_j^*(x_j-1,u_j), G_j^*(x_j,u_j)]$ is strictly larger than $(G_i^*(x_i-1,u_i), G_i^*(x_i,u_i)]$ but $w_i \geqslant w_j$. This contradicts the condition that whenever $G_i^*(x_i,u_i) \leqslant G_j^*(x_j-1,u_j)$, then $w_i < w_j$. □

We present a Bernstein-von Mises theorem for the fiducial distribution associated with the empirical distribution function. This result can be viewed either as a special case (without censoring) of [4], or as a particular case of the exchangeably weighted bootstrap in [39].

Corollary A.1. *Assume the conditions of Theorem* 3.2. *We have*

$$n^{1/2}\{\widetilde{F}(\cdot) - \widehat{F}_n(\cdot)\} \to \{1-F(\cdot)\}B(\gamma(\cdot))$$

in distribution on Skorokhod space $\mathcal{D}[0,1]$ *in probability, where* B *is the Brownian motion,*

$$\gamma(t) = \int_0^t \frac{f(s)}{[1-F(s)]^2}ds = \frac{F(t)}{1-F(t)},$$

$\widehat{F}_n$ *is the oracle empirical distribution function defined in* (3.2)*, and* $\widetilde{F}$ *is the oracle fiducial distribution based on the unobserved* $P_1,\ldots,P_n$*,*

$$\widetilde{F}(s) = \sum_{i=0}^{n} I[P_{(i)} \leqslant s < P_{(i+1)}]W_{(i)}^*, \tag{A.1}$$

where $P_{(0)} \equiv 0$*,* $P_{(n+1)} \equiv 1$*,* $W_{(i)}^*$ *are uniform order statistics, and* $W_{(0)}^* \equiv 0$*.*

Proof. Notice (A.1) is the lower fiducial distribution F^L in [4] when there is no censoring. By [4, Theorem 2], we essentially need to check its assumptions 1–3. Assumption 1 is satisfied with $\pi(p) = 1-F(p)$; Assumption 2 is satisfied as we assume that the true CDF is absolutely continuous; Assumption 3 is satisfied as

$$\int_0^p \frac{g_n(s)}{\sum_{i=1}^n I(P_i \geqslant s)} d\left[\sum_{i=1}^n I(P_i \leqslant s)\right] \to \int_0^p \frac{f(s)}{[1-F(s)]^2}ds$$

for any p such that $1-F(p) > 0$ and any sequence of functions $g_n \to \frac{1}{1-F}$ uniformly. □

Lemma A.2. *Assume the conditions of Theorem* 3.2. *Suppose that* $X_i \sim \mathrm{Bin}(m_i,P_i)$*, where* P_i*'s are the unobserved i.i.d. random variables with the distribution function* F*. We have*

$$\min_{i,j\in\{0,\ldots,n\}} \left\{\frac{X_i}{m_i} - \frac{X_j}{m_j}\right\} \succeq O\left(\frac{1}{n^2(\log n)\sqrt{\epsilon/2}}\right) \tag{A.2}$$

for any $\epsilon > 0$ *with a probability converging to* 1*.*

Proof. We first show that the unobserved $P_i = F^{-1}(W_i)$ are well separated. Straightforward calculation with uniform order statistics shows that

$$\mathrm{pr}\bigg(\min_{i\in\{0,\ldots,n\}}\{W_{(i+1)} - W_{(i)}\} > \frac{t}{n(n+1)}\bigg) \geqslant \bigg(1 - \frac{t}{n}\bigg)^n$$

for any $t > 0$, where $W_{(0)} \equiv 0$ and $W_{(n+1)} \equiv 1$. Therefore,

$$\min_{i\in\{0,\ldots,n\}}\{P_{(i+1)} - P_{(i)}\} \succeq O\bigg(\frac{1}{n^2(\log n)\sqrt{\epsilon/2}}\bigg)$$

with a probability converging to 1, where $P_{(0)} \equiv 0$ and $P_{(n+1)} \equiv 1$. Furthermore, by the following Bernstein inequality for $X \sim \mathrm{Bin}(m, P)$, we see that

$$\mathrm{pr}(|X - mP| \geqslant t) \leqslant 2\exp\bigg\{-\frac{t^2}{2[mP(1-P) + t/3]}\bigg\},$$

where $t > 0$. Taking $t \sim O(\frac{m}{n^2(\log n)\sqrt{\epsilon/2}})$, we have

$$\min_{i,j\in\{0,\ldots,n\}}\bigg\{\frac{X_i}{m_i} - \frac{X_j}{m_j}\bigg\} \succeq O\bigg(\frac{1}{n^2(\log n)\sqrt{\epsilon/2}}\bigg)$$

with a probability converging to 1 because $\lim_{n\to\infty} n^4(\log n)^{1+\epsilon}/m_i = 0$ for any $m = m_i$, given by Assumption 3.1. □

Lemma A.3. *Given data satisfying* (A.2)*, if* $(\boldsymbol{U}^\star, \boldsymbol{W}^\star)$ *are uniformly distributed on the set* $\{(\boldsymbol{u}^\star, \boldsymbol{w}^\star) : Q_{\boldsymbol{x}}(\boldsymbol{u}^\star, \boldsymbol{w}^\star) \neq \emptyset\}$*, we have*

$$\mathrm{pr}\{(G_i^*(x_i - 1, U_i^\star), G_i^*(x_i, U_i^\star)] \cap (G_j^*(x_j - 1, U_j^\star), G_j^*(x_j, U_j^\star)] \neq \emptyset \text{ for some } i \neq j\} \to 0. \tag{A.3}$$

Proof. Let $\widetilde{U}_i$ be i.i.d. $U(0,1)$, and define

$$L_i = G_i^*(x_i - 1, \widetilde{U}_i) \sim \mathrm{Beta}(x_i, m_i - x_i + 1)$$

and

$$R_i = G_i^*(x_i, \widetilde{U}_i) \sim \mathrm{Beta}(x_i + 1, m_i - x_i).$$

Recall that the proposed Algorithm 1 can be regarded as an importance sampling method in the following way: if there is one k-intersection (i.e., $k+1$ intervals share one common area) in $\{(L_i, R_i], i = 1, \ldots, n\}$, the corresponding W^* have $(k+1)!$ possible permutations. So we essentially need to show that

$$\sum_{k=1}^{n-1}(k+1)!\, q_k \to 0$$

as $n \to 0$, where q_k is defined as the probability of $\{(L_i, R_i], i = 1, \ldots, n\}$ having one or more k-intersections, and $W_{i_1}^* < \cdots < W_{i_{k+1}}^*$ (i.e., $i_1, \ldots, i_{k+1}$ are indices of intervals that intersect).

We start with a one-intersection between the i-th and j-th intervals, i.e., $(L_i, R_i) \cap (L_j, R_j) \neq \emptyset$, which is equivalent to $L_1 \leqslant R_2$ and $L_2 \leqslant R_1$. Without loss of generality, we focus on $L_1 \leqslant R_2$ as $\mathrm{pr}(A_1 \cap A_2) \leqslant \min\{\mathrm{pr}(A_1), \mathrm{pr}(A_2)\}$. For a random variable $Y \sim \mathrm{Beta}(\alpha, \beta)$, let $\mu = E[Y] = \frac{\alpha}{\alpha+\beta}$. By [30, Theorem 2.1], Y is sub-Gaussian with the variance proxy parameter $\Sigma \equiv \frac{1}{4(\alpha+\beta+1)}$. Hence by the definition of the sub-Gaussian random variable, for any $c, t \in \mathbb{R}$, $\mathrm{pr}(Y - \mu \geqslant t) \leqslant \exp\{-\frac{t^2}{2\Sigma}\}$.

Note that

$$\mathrm{pr}(L_1 \leqslant R_2) = \mathrm{pr}(R_2 - L_1 \geqslant 0), \tag{A.4}$$

where L_1 is sub-Gaussian with mean $x_1/(m_1+1)$ and variance $\Sigma_1 = 1/(m_1+2)$, and R_2 is sub-Gaussian with mean $(x_2+1)/(m_2+1)$ and variance $\Sigma_2 = 1/(m_2+2)$. Therefore, (A.4) equals $\mathrm{pr}(Z \geqslant x_1/(m_1+1) - (x_2+1)/(m_2+1))$, where Z is sub-Gaussian with mean 0 and variance $\Sigma_Z = 1/(m_1+2) + 1/(m_2+2)$.

Then we further have that (A.4) equals

$$\mathrm{pr}\{Z \geqslant x_1/(m_1+1) - (x_2+1)/(m_2+1)\} \leqslant \exp\left\{ -\frac{T^2}{2\Sigma_Z} \right\} \leqslant \exp\{-\tilde{m}T^2\},$$

where $\tilde{m} = \min_{i=1,\ldots,n} m_i$ and $T = x_1/(m_1+1) - (x_2+1)/(m_2+1)$.

Thus, the probability of a single intersection $2!\, q_1$ is bounded by

$$2! \binom{n}{2} \exp[-\tilde{m}T^2] \preceq \exp\{c_1 \log n - c_2 (\log n)^{1+\epsilon/2}\}$$

for any $\epsilon > 0$ and some constants c_1 and c_2. We note that the coefficient $\binom{n}{2}$ refers to the number of possible pairs.

Next, we consider the case of k-intersections. We only need to consider the two intervals corresponding to the two farthest P_i among the k intervals. Thus, the probability of existing k-intersections $(k+1)!\, q_k$ is bounded by

$$\begin{aligned}(k+1)! \binom{n}{k} \exp\left\{ -\tilde{m}\left[O\left(\frac{k}{n^2 (\log n)^{\sqrt{\epsilon/2}}} \right) \right]^2 \right\} \\ \preceq \exp\{c_1 (k \log n + k \log k) - c_2 k^2 (\log n)^{1+\epsilon/2}\}\end{aligned}$$

for any $\epsilon > 0$ and some constants c_1 and c_2. Because $\sum_{k=1}^{n-1} (k+1)!\, q_k \to 0$, we conclude that (A.3) holds. □

Proof of Theorem 3.2. Recall that the data-generating equation is

$$X_i = G_i^{-1}(U_i, P_i), \quad P_i = F^{-1}(W_i), \tag{A.5}$$

where G_i is the CDF of the binomial distribution.

By Corollary A.1, we define

$$\widetilde{F}^L(s) \equiv \sum_{i=0}^{n} I[G^U_{(i)} \leqslant s < G^U_{(i+1)}] W^*_{(i)}$$

and

$$\widetilde{F}^U(s) \equiv \sum_{i=0}^{n} I[G^L_{(i)} \leqslant s < G^L_{(i+1)}] W^*_{(i+1)},$$

where $G^U_{(1)}, \ldots, G^U_{(n)}$ are order statistics of $\{G^*_i(x_i, U^\star_i), i = 1, \ldots, n\}$, $G^L_{(1)}, \ldots, G^L_{(n)}$ are order statistics of $\{G^*_i(x_i - 1, U^\star_i), i = 1, \ldots, n\}$, $G^U_{(0)} \equiv 0$, $G^U_{(n+1)} \equiv 1$, $G^L_{(0)} \equiv 0$, $G^L_{(n+1)} \equiv 1$, and $W^*_{(n+1)} \equiv 1$. Note that $\widetilde{F}^L$ and $\widetilde{F}^U$ can be regarded as lower and upper bounds of F^L and F^U, respectively.

In order to obtain $n^{1/2}\{F^L(\cdot) - \widehat{F}_n(\cdot)\} \to \{1 - F(\cdot)\} B(\gamma(\cdot))$ in distribution in probability, it is enough to show that $\sup_s n^{1/2} |\widetilde{F}(s) - F^L(s)| \to 0$ in probability, which is implied by

$$\sup_s n^{1/2} \{\widetilde{F}^U(s) - \widetilde{F}^L(s)\} \to 0 \tag{A.6}$$

in probability. In order to show (A.6), one essentially needs to show that for any $\epsilon > 0$,

$$\mathrm{pr}\Big(\sup_s n^{1/2} \{\widetilde{F}^U(s) - \widetilde{F}^L(s)\} > \epsilon \Big) \to 0.$$

By Lemma A.3, there is no intersection between

$$(G^*_i(x_i - 1, U^\star_i), G^*_i(x_i, U^\star_i)], \quad i = 1, \ldots, n$$

with a probability converging to 1.

Thus, we have

$$\begin{aligned}\mathrm{pr}\Big(\sup_s n^{1/2}\{\widetilde{F}^U(s)-\widetilde{F}^L(s)\}>\epsilon\Big) &\leqslant \sum_{i=0}^{n}\mathrm{pr}\bigg(W^*_{(i+1)}-W^*_{(i)}>\frac{\epsilon}{n^{1/2}}\bigg)\\ &=(n+1)\mathrm{pr}\bigg(\mathrm{Beta}(1,n)>\frac{\epsilon}{n^{1/2}}\bigg)\\ &=(n+1)\bigg(1-\frac{\epsilon}{n^{1/2}}\bigg)^n\to 0.\end{aligned}$$

Therefore, we have

$$n^{1/2}\{F^L(\cdot)-\widehat{F}_n(\cdot)\}\to\{1-F(\cdot)\}B(\gamma(\cdot))$$

in distribution in probability. Note that for any $t<s$,

$$\mathrm{cov}[\{1-F(s)\}B(\gamma(s)),\{1-F(t)\}B(\gamma(t))]=\gamma(t)\{1-F(s)\}\{1-F(t)\}=F(t)\{1-F(s)\},$$

which completes the proof. □

Appendix B Remarks on binomial and Poisson data

In the following two theorems for binomial and Poisson data, respectively, we show implications of the fact that $\mathrm{pr}(F\in Q_x(U^*,W^*))$ is proportional to the nonparametric likelihood function. In particular, maximizing the scaled fiducial probability in its limit provides exactly the underlying true CDF.

Theorem B.1. *Suppose that $\Theta_i\equiv P_i$ and $X_i\mid P_i\sim\mathrm{Bin}(m,P_i)$. Maximizing*

$$\lim_{n\to\infty}[c_n\mathrm{pr}(F\in Q_{\boldsymbol{x}}(\boldsymbol{U}^\star,\boldsymbol{W}^\star))]^{1/n}$$

leads to a CDF matching the first m moments of the true $F(p)$, where c_n is the normalizing constant.

Proof. Recall the data-generating equation (A.5), where G_i is the CDF of the binomial distribution. The fiducial probability is

$$\begin{aligned}\mathrm{pr}(F\in Q_{\boldsymbol{x}}(\boldsymbol{U}^\star,\boldsymbol{W}^\star))&\propto\prod_{i=1}^{n}\bigg[\int_0^1\binom{m}{x_i}p^{x_i}(1-p)^{m-x_i}dF(p)\bigg]\\ &=\exp\bigg\{\sum_{i=1}^{n}\log\bigg(E_F\bigg[\binom{m}{x_i}p^{x_i}(1-p)^{m-x_i}\bigg]\bigg)\bigg\}\\ &=\exp\bigg\{n_m\log(E_F[P^m])+n_{m-1}\log\bigg(E_F\bigg[\binom{m}{m-1}P^{m-1}(1-P)\bigg]\bigg)+\cdots\\ &\quad+n_0\log(E_F[(1-P)^m])\bigg\},\end{aligned}$$

where n_k is the number of samples with $X_i=k$, and E_F refers to the expectation with respect to F that is evaluated. Thus, as n goes to infinity,

$$\begin{aligned}&[c_n\mathrm{pr}(F\in Q_{\boldsymbol{x}}(\boldsymbol{U}^\star,\boldsymbol{W}^\star))]^{1/n}\\ &\quad\to\exp\bigg\{E[P^m]\log(E_F[P^m])+E\bigg[\binom{m}{m-1}P^{m-1}(1-P)\bigg]\log\bigg(E_F\bigg[\binom{m}{m-1}P^{m-1}(1-P)\bigg]\bigg)\\ &\qquad+\cdots+E[(1-P)^m]\log(E_F[(1-P)^m])\bigg\},\end{aligned}\tag{B.1}$$

where c_n is the normalizing constant, and E refers to the expectation with respect to the true distribution function that generates data. By the method of Lagrange multipliers,

$$H(x)=\sum_i y_i\log x_i\quad\text{subject to }\sum_i x_i=1,\quad\sum_i y_i=1$$

is maximized with respect to x by setting $x_k = y_k$, $k \in \mathbb{R}^+$. Maximizing (B.1) gives

$$\begin{aligned} &E[P^m] = E_F[P^m], \\ &E[P^{m-1}(1-P)] = E_F[P^{m-1}(1-P)], \\ &\qquad\vdots \\ &E[(1-P)^m] = E_F[(1-P)^m]. \end{aligned}$$

The above equations are restrictions on the first m moments of P, which completes the proof. □

Theorem B.2. *Suppose that* $\Theta_i \equiv \Lambda_i$ *and* $X_i \mid \Lambda_i \sim \mathrm{Poi}(\Lambda_i)$. *Maximizing*

$$\lim_{n\to\infty} [c_n \mathrm{pr}(F \in Q_{\boldsymbol{x}}(\boldsymbol{U}^\star, \boldsymbol{W}^\star))]^{1/n}$$

leads to true $F(\lambda)$ *almost surely, where* c_n *is the normalizing constant.*

Proof. Recall the data-generating equation $X_i = G_i^{-1}(U_i, \Lambda_i)$, $\Lambda_i = F^{-1}(W_i)$, where G_i is the CDF of the Poisson distribution. The fiducial probability is

$$\begin{aligned} &\mathrm{pr}(F \in Q_{\boldsymbol{x}}(\boldsymbol{U}^\star, \boldsymbol{W}^\star)) \\ &\quad\propto \prod_{i=1}^n \int_0^\infty \frac{\lambda^{x_i}\exp\{-\lambda\}}{x_i!} dF(\lambda) \\ &\quad= \left\{\frac{E_F[\exp(-\Lambda)]}{0!}\right\}^{n_0} \left\{\frac{E_F[\Lambda\exp(-\Lambda)]}{1!}\right\}^{n_1} \cdots \left\{\frac{E_F[\Lambda^k\exp(-\Lambda)]}{k!}\right\}^{n_k} \cdots \\ &\quad= \exp\left\{n\left[\frac{n_0}{n}\log\frac{E_F[\exp(-\Lambda)]}{0!} + \frac{n_1}{n}\log\frac{E_F[\Lambda\exp(-\Lambda)]}{1!} + \cdots + \frac{n_k}{n}\log\frac{E_F[\Lambda^k\exp(-\Lambda)]}{k!} + \cdots\right]\right\}, \end{aligned}$$

where n_k is the number of samples with $X_i = k$, and E_F refers to the expectation with respect to F that is evaluated. Thus, as n goes to infinity,

$$\begin{aligned} &[c_n \mathrm{pr}(F \in Q_{\boldsymbol{x}}(\boldsymbol{U}^\star, \boldsymbol{W}^\star))]^{1/n} \\ &\quad\to \exp\left\{\frac{E[\exp(-\Lambda)]}{0!}\log\frac{E_F[\exp(-\Lambda)]}{0!} + \frac{E[\Lambda\exp(-\Lambda)]}{1!}\log\frac{E_F[\Lambda\exp(-\Lambda)]}{1!} + \cdots\right. \\ &\qquad\left. + \frac{E[\Lambda^k\exp(-\Lambda)]}{k!}\log\frac{E_F[\Lambda^k\exp(-\Lambda)]}{k!} + \cdots\right\}, \end{aligned} \tag{B.2}$$

where c_n is the normalizing constant, and E refers to the expectation with respect to the true distribution function that generates data. By the method of Lagrange multipliers, maximizing (B.2) gives

$$\frac{E[\Lambda^k\exp(-\Lambda)]}{k!} = \frac{E_F[\Lambda^k\exp(-\Lambda)]}{k!}, \quad k \in \mathbb{R}^+.$$

The above equations essentially impose restrictions on all the derivatives of the Laplace transform at 1. The fact that the Laplace transform is analytic in the region of absolute convergence implies the uniqueness of the distribution, which completes the proof. □