JOURNAL OF BIOPHARMACEUTICAL STATISTICS
https://doi.org/10.1080/10543406.2021.1946692

Meta inference of heterogeneous data streams

Saeid Amiri

The Neuro (Montreal Neurological Institute-Hospital), McGill University, Montréal, Quebec, Canada

ABSTRACT
This work aims to explore the meta inference of mean, where data are being collected from different studies. Providing an accurate estimate of the mean from separate studies is a central aspect of the practical sciences. The meta inference of means and its statistical inference is so prevalent that it has received much attention. The problem is well recognized as pertinent to carrying out the meta inference of mean, and its statistical inference has got a lot of attention. This paper intends to explore the existing techniques for dealing with this issue, propose and study alternative methods to improve the estimation of a common mean, and review the theoretical inference. This paper uses numerical investigations that continue to provide good results to evaluate the proposed methods.

ARTICLE HISTORY
Received 03 June 2020
Accepted 14 June 2021

KEYWORDS
Bootstrap method; maximum likelihood estimator; meta inference; nonparametric method

1. Introduction

In recent years, more and more data have become available continuously from different sources, e.g., health apps in smart watches that record the health and fitness information. Having data from the heterogeneous data stream, different sources/studies, need an appropriate inference. Method of drawing inference of multiple studies is often referred to as *meta inference*, an approach to estimate that typically performs better than estimation from any single data set. Also, in many practical studies, the data are collected from different studies to draw a statistical inference due to insufficient sample size. Estimations from this data are more realistic, and the resulting heterogeneity emerges due to the disparate conditions in which data were collected. An optimal estimate can more likely be found with heterogeneous populations; this intriguing idea from Graybill and Deal (1959) has received serious attention in research from past decades as a way to overcome the difficulties of combining several studies to provide more informative results. For a brief review and latest developments on this common normal mean problem, see Chang and Pal (2008), Hartung et al. (2011), Ma et al. (2011), and Kim (2013) as well as the references therein.

Practitioners in scientific disciplines, more specifically in medical sciences, are often exposed to such data where they are collected from multiple studies/sources. Dealing with such data needs accurate statistical inferences; the problems of accurate estimate and its statistical inference have been studied for four decades but still of interest because of its applicabilities to real-life data. The essential of finding an appropriate inference of medical research has been explored (see Brockwell and Gordon (2001) and references therein). Recent technical research on the combining means can be found in Chang et al. (2012), among others. We refer the reader to a monograph by Keller et al. (2004, which provides invaluable information on the meta inference, its many variants, theories, and applications. The estimate of mean is computed under the assumption that the samples collected from independent studies are from normal populations with a common mean but possibly unequal variances. An important consideration may be used to estimate or construct a confidence interval for the common mean of these populations in such cases. This paper is devoted to illustrating different

CONTACT Saeid Amiri saeid.amiri1@gmail.com 3801 Rue University, Montréal, QC H3A 2B4

methods for estimating the mean and presenting a nonparametric method that provides an accurate estimate. Its statistical inference is also studied.

We introduce some necessary notations and symbols that will be used throughout the paper. Suppose a total number of n units are available from K different studies that are to be measured from the underlying populations on the variable of interest. Let n_r be the number of measurements on the rth sample, $r = 1, \ldots, K$, such that $n = \sum_{r=1}^{K} n_r$. Let $X_{r,j}$ denote the measurement on the jth measured unit in the rth sample. This results in samples of size n_r from the underlying populations as

$$\{X_{r,j}; r = 1, \ldots, K, j = 1, \ldots, n_r\}.$$

It is worth mentioning that, in our setup designs, the observations are independent but not identical. We can represent the structure of data as

$$\mathcal{X}_1 = \{X_{1,1}, X_{1,2}, \ldots, X_{1,n_1}\} \overset{i.i.d.}{\sim} F_1(x; \mu, \sigma_1),$$

$$\mathcal{X}_2 = \{X_{2,1}, X_{2,2}, \ldots, X_{2,n_2}\} \overset{i.i.d.}{\sim} F_2(x; \mu, \sigma_2),$$

$$\vdots$$

$$\mathcal{X}_K = \{X_{K,1}, X_{K,2}, \ldots, X_{K,n_K}\} \overset{i.i.d.}{\sim} F_K(x; \mu, \sigma_K),$$

where $F_r(x; \mu, \sigma_r)$ is the distribution function (df) of the rth population, and we assume the populations with the same mean but different variances or heterogeneity. This design is more general than the one considered in Brockwell and Gordon (2001), where they constrained the design under the normality of observations. We attempted to cover several issues: finding an accurate nonparametric method for estimating the common mean, establishing a general inferential basis for the common mean, and using an asymptotic approach because the weight is a function of observations.

This paper investigates the methods for estimating the mean of heterogeneous samples. To this end, we offer a new nonparametric method for estimating the mean that, unlike the existing approaches, considers the heterogeneity of other samples to estimate the weights. We also explore the weighted mean via the square of deviance from the grand mean. Presenting any point estimation method without appropriate statistical inference is untenable; hence, we present the statistical inference via the bootstrap method, a nonparametric approach. From a theoretical point of view, we provide an asymptotic inference of the proposed methods, so this paper aims to shed light on a range of issues related to statistical inference of a common mean in a general setting.

To achieve this aim, the remainder of this paper is organized as follows. Section 2 explores the direct combination of multiple samples. Section 3 discusses the weighted estimates of mean, which includes the Graybill–Deal estimate, Maximum Likelihood method, and new approaches to estimate the mean in a similar way as the Graybill–Deal estimate. We also present theoretical results that suggest the proposed methods should work well under some generality. Section 4 explores the statistical inference via the bootstrap method. Section 5 examines the efficiency of the discussed methods through numerical investigation. We provide a series of examples, both simulated and real data, that verify the performance of the discussed techniques.

2. Combining samples

To provide insight into the nature of the naive estimator, here is a review of the basic approaches to combining samples; the classified edf of data is defined as

$$\widehat{F}_{n_K}(t) = \frac{1}{K}\sum_{r=1}^{K}\frac{1}{n_r}\sum_{j=1}^{n_r} I(X_{r,j} \leq t) = \frac{1}{K}\sum_{r=1}^{K}\widehat{F}_r(t).$$

If $n_r \to \infty$, $r = 1, \ldots, K$, using the Glivenko–Cantelli Theorem, we can show

$$\sup_{t\in R}\{\hat{F}_r(t) - F_r(t)\} \longrightarrow 0 \text{a.s.} \forall t, n_r \to \infty,$$

hence,

$$\sup_{t\in\mathbb{R}}\{\widehat{F}_{n_K}(t) - F_c(t)\} \to 0 \text{a.s.} \forall t, n_r \to \infty,$$

where

$$F_c(t) = \frac{1}{K}\sum_{r=1}^{K} F_r(t). \tag{1}$$

The other approach is to define the pooled *edf*,

$$\widehat{F}_{q_n}(t) = \frac{1}{n}\sum_{r=1}^{K}\sum_{j=1}^{n_r} I(X_{r,j} \leq t) = \sum_{r=1}^{K} q_{n_r}\widehat{F}_r(t),$$

where $q_{n_r} = n_r/n$ and

$$F_q(t) = \sum_{r=1}^{K} q_{n_r} F_r(t), \tag{2}$$

where under the equal-size samples, $F_c(.)$ and $F_q(.)$ in Equations (1) and (2) are the same. The sample mean of each sample and its variance can be represented as

$$\bar{X}_r = \frac{1}{n_r}\sum_{j=1}^{n_r}\bar{X}_{r,j},$$

$$V(\bar{X}_r) = \frac{\sigma^2(F_r)}{n_r}.$$

The classified estimate mean, *CM*, can be represented as

$$\widehat{\mu}_{CM} = \frac{1}{K}\sum_{r=1}^{K}\bar{X}_r = \frac{1}{K}\sum_{r=1}^{K}\frac{1}{n_r}\sum_{j=1}^{n_r}\bar{X}_{r,j},$$

where the expected value $E(.)$ is

$$E(\widehat{\mu}_{CM}) = \frac{1}{K}\sum_{r=1}^{K} E(\bar{X}_r) = \mu,$$

and due to the independence of the samples, the variance is

$$V(\widehat{\mu}_{CM}) = \frac{1}{K^2}\sum_{r=1}^{K} V(\bar{X}_r),$$

where

$$V(\bar{X}_r) = \frac{\sigma^2(F_r)}{n_r}, r = 1, \ldots, K,$$

where $V(.)$ stands for the variance. The following proposition provides an asymptotic result for the classified mean.

Proposition 1. *Suppose the rth sample is generated from $F_r(.;\mu, \sigma^2(F_r))$ with $\int x^2 dF_r(x) < \infty$ and $\widehat{F}_r(.)$ is the edf of the rth sample. If $\vartheta_r = (\theta(\widehat{F}_r) - \theta(F_r))$, then $(\vartheta_1, \ldots, \vartheta_K)$ converges in distribution to a multivariate normal distribution with mean vector zero and covariance matrix $diag(\sigma^2(F_1)/n_1, \ldots, \sigma^2(F_k)/n_K)$ where, $\sigma^2(F_r) = \int (X - \mu_r)^2 dF_r(x)$ and $\mu_r = \int x dF_r(x)$.*

This proposition suggests that the statistic is held as

$$\frac{\frac{1}{K}\sum_{r=1}^{k} \bar{X}_r - \mu}{S_c} \to N(0,1),$$

where

$$S_c^2 = \frac{1}{K^2}\sum_{r=1}^{K}\frac{\sigma^2(\widehat{F}_r)}{n_r}.$$

Using the Central Limit Theorem can approximate a confidence interval, where

$$P(\mu \in (\frac{1}{K}\sum_{r=1}^{K}\bar{X}_r - z_{\alpha/2}S_c, \frac{1}{K}\sum_{r=1}^{K}\bar{X}_r + z_{\alpha/2}S_c)) \approx 1 - \alpha.$$

Under the pooled samples, PM, the estimator is as follows:

$$\widehat{\mu}_{PM} = \frac{1}{n}\sum_{r=1}^{K}\sum_{j=1}^{n_r} X_{r,j}. \quad (3)$$

This estimate is unbiased and

$$\frac{\frac{1}{n}\sum_{r=1}^{K}\sum_{j=1}^{n_r} X_{r,j} - \mu}{S_p} \to N(0,1),$$

where

$$S_p^2 = \frac{1}{n^2}\sum_{r=1}^{K} n_r \sigma^2(\widehat{F}_r).$$

It is of interest to compare the classified and pooled estimators, so we compare these estimators for $K = 2$. By simple algebra, we can show that under $\sigma_1 = \sigma_2$,

$$V(\widehat{\mu}_{CM}) \geq V(\widehat{\mu}_{PM}),$$

but when holding condition

$$\frac{\sigma_1^2}{\sigma_2^2} \geq \frac{n^2 n_1 - 4n_2^2 n_1}{4n_1^2 n_2 - n_2 n^2},$$

$V(\widehat{\mu}_{PM}) \geq V(\widehat{\mu}_{CM})$; hence, the performance of these estimators depends on the parameters which might not be of interest in practical. For example, let us consider $n_2 = 2n_1$, in which then under $\sigma_2 = 2\sigma_1$, we have $V(\widehat{\mu}_{CM}) < V(\widehat{\mu}_{PM})$, but inversely under $\sigma_1 = 2\sigma_2$, $V(\widehat{\mu}_{CM}) > V(\widehat{\mu}_{PM})$ is held; hence, one needs a table of values to find which method is good for which n's and σ's. The discussion under $K>2$ may not be trivial; we numerically examined their performances in Section 5, we presented a better estimate than these CM and PM; therefore, their theoretical comparison under $K>2$ is not persuaded anymore.

3. Weighted mean

In general, the classified mean can be expressed as follows:

$$\widehat{\mu}_w = \sum_{r=1}^{K} w_r \bar{X}_r,$$

where $\sum_{r=1}^{K} w_r = 1$, which is a linear combination of the sample means. Let assume w_rs are independent of observations, then

$$E(\widehat{\mu}_w) = \sum_{r=1}^{K} w_r E(\bar{X}_r) = \mu,$$

where

$$V(\widehat{\mu}_w) = \sum_{r=1}^{K} w_r^2 V(\bar{X}_r).$$

The $\widehat{\mu}_{CM}$ estimate is a particular case obtained under $w_r = \frac{1}{K}$. The most appealing estimator is known as the Graybill–Deal estimator (GDE, see [Graybill and Deal 1959)), given as

$$\widehat{w}_{r,GDE} = \frac{(n_r/s_r^2)}{\sum_{r=1}^{K}(n_r/s_r^2)}, r = 1,\ldots,K, \tag{4}$$

where $\widehat{w}_r$ is a function of observations, and we denote the estimator as $\widehat{\mu}_{GDE}$, where $\widehat{\mu}_{GDE} = \sum_{r=1}^{K} \widehat{w}_{r,GDE}\bar{X}_r$.

The other estimator is the maximum likelihood estimator (MLE), which can be obtained by assuming the normality of parent distribution. The joint log-likelihood function of $\mu, \sigma_1, \ldots, \sigma_K$ based on the samples can be written as

$$L(\mu, \sigma_1, \ldots, \sigma_K; X) = \log(\prod_{r=1}^{K}\prod_{j=1}^{n_r} f(X_{r,j}; \mu, \sigma_r))$$

$$= -\sum_{r=1}^{K}(\frac{n_r}{2}(\log(2\pi) + \log(\sigma_r^2)) + \sum_{j=1}^{n_r}\frac{(X_{r,j}-\mu)^2}{2\sigma_r^2})$$

$$= -\sum_{r=1}^{K}\frac{n_r}{2}(\log(2\pi) + \log(\sigma_r^2) + \frac{s_r^2}{\sigma_r^2} + \frac{(\bar{X}_r-\mu)^2}{\sigma_r^2}). \tag{5}$$

There is no close form for the MLEs, but it can be found numerically. Assuming there is a unique parent distribution (μ, σ_0), and the variabilities based on situations in which data collected, the variance can be represented as $\sigma_r^2 = c_r^2\sigma_0^2$. Equation (5) can be written as the following, which enables us to estimate these parameters:

$$L(\mu, \sigma_1, \ldots, \sigma_K; \mathbf{X}) = -\sum_{r=1}^{K} \frac{n_r}{2} (\log(2\pi) + \log(c_r^2\sigma_0^2) + \frac{s_r^2}{c_r^2\sigma_0^2} + \frac{(\bar{X}_r - \mu)^2}{c_r^2\sigma_0^2}). \tag{6}$$

Let us denote the MLE estimate as

$$\widehat{\mu}_{ML} = \sum_{r=1}^{K} \widehat{w}_{r,ML}\bar{X}_r, \tag{7}$$

where

$$\widehat{w}_{r,ML} = \frac{(n_r/\widehat{\sigma}_r^2)}{\sum_{r=1}^{2} (n_r/\widehat{\sigma}_r^2)}, r = 1, \ldots, K. \tag{8}$$

Under $K = 2$, it can be shown that

$$\widehat{\sigma}_1^2 = s_1^2 + (\frac{n_2\widehat{\sigma}_1^2}{n_2\widehat{\sigma}_1^2 + n_1\widehat{\sigma}_2^2})^2(\bar{x}_1 - \bar{x}_2)^2.$$

The estimate of $\widehat{\sigma}_2^2$ obtained by exchanging the role of the subscripts (see Kim (2013)). Holding the normality of parent distribution is a firm assumption and makes it less interesting, although the violation should not cause a problem for large sample sizes. Both GDE and MLE are commonly used "inverse variance" estimator with different ways of estimating within-study variance.

Here we choose another approach to estimate the common mean: the weights in the *GDE* created using the variabilities inside the corresponding samples. However, it might be fruitful to use a method that emphasizes the variabilities of samples rather than *r*th sample. To build such an estimator, we define the *r*th weight in terms of the variabilities of samples excluding the *r*th sample as

$$\widehat{w}_{r,SV} = \frac{\sum_{j\neq r} (s_j^2/n_j)}{\sum_{i=1}^{K} \sum_{j\neq i} (s_j^2/n_j)}, r = 1, \ldots, K, \tag{9}$$

where $\sum_{r=1}^{K} \widehat{w}_{r,SV} = 1$. Obviously, unlike the *r*th weight in *GDE*, here the variabilities of samples other than the *r*th sample are used in the nominator of the *r*th weight. This estimate is referred to as *Shared Variability without the rth sample estimator* (*SV*) in the rest of this work. Under $K = 2$, it is trivial to show $\widehat{w}_{r,SV} = \widehat{w}_{r,GDE}, r = 1, 2$, but for $K > 2$ they perform differently.

The existing methods use the variabilities inside samples; here we consider building an estimator using the variability from a grand mean, potentially running a double-weighted approach. By estimating $\widehat{w}_{r,GDE}$, (4), calculating $\widehat{\mu}_{GDE}$, and then estimating the square deviance of samples from $\widehat{\mu}_{GDE}$,

$$\mathcal{D}_r^2 = \frac{1}{n_r - 1} \sum_{j=1}^{n_r} (X_{r,j} - \widehat{\mu}_{GDE})^2,$$

the weight is obtained as follows:

$$\widehat{w}_{r,DGDE} = \frac{(n_r/D_r^2)}{\sum_{r=1}^{K} (n_r/D_r^2)}, r = 1, \ldots, K. \tag{10}$$

It is referred to as Double *GDE* (*DGDE*) in the rest of this work. A key feature of this method is that it can be considered as a fully nonparametric approach and simple to calculate. The same discussion can be done for $\widehat{\mu}_{SV}$, and the estimate is denoted as *DSV*.

It can be shown that the weights obtained via the reciprocal of variance provide optimal estimators, in some respects, and better performance than the weighted means using $\sqrt{s_r^2}$ or s_r^3. Our numerical analysis shows the weighted means using $\sqrt{s_r^2}$ or s_r^3 perform poorly; hence, we do not discuss them here. Under holding the normality of data and using the MLE, it is trivial to show the optimality of using s_r^2. Let consider the mean squared error (MSE) of an estimator $\widehat{\theta}$,

$$MSE(\widehat{\theta}) = (E(\widehat{\theta}) - \theta)^2 + V(\widehat{\theta}),$$

for the given weighted mean $E(\widehat{\mu}) = \sum_{r=1}^{K} w_r \mu_r$,

$$MSE(\widehat{\mu}) = (E(\widehat{\mu}) - \mu)^2 + V(\widehat{\mu}) = (\sum_{r=1}^{K} w_r \mu_r - \mu)^2 + \sum_{r=1}^{K} w_r^2 \frac{\sigma_r^2}{n_r}.$$

Using the derivative of $MSE(\widehat{\mu})$, one can find the minimum:

$$\frac{\partial MSE(\widehat{\mu})}{\partial w_r} = 2\mu_r(\sum_{r=1}^{K} w_r \mu_r - \mu) + 2w_r \frac{\sigma_r^2}{n_r} = 0.$$

Let us denote $\delta_0 = \mu - \sum_{r=1}^{K} w_r \mu_r$ that is the amount of sample bias, where $\delta_0 \neq 0$. The minimum MSE occurs at the following point:

$$w_{r0} = \delta_0 n_r \frac{\mu_r}{\sigma_r^2}.$$

Therefore, the naive estimate (plug-in) of δ_0 and w_{r0} are

$$\widehat{\delta}_0 = \widehat{\mu} - \sum_{r=1}^{K} \widehat{w}_r \bar{X}_r,$$

$$\widehat{w}_{r0} = \widehat{\delta}_0 n_r \frac{\bar{X}_r}{s_r^2}.$$

For the r sample, the ratio of sample mean and variance ($\frac{\bar{X}_r}{s_r^2}$) plays a role in the estimate $\widehat{w}_{r0}$. Obviously, it can be written in terms of the coefficient of variation, that is $\frac{1}{\bar{X}_r \widehat{\gamma}_r}$, where $\widehat{\gamma}_r$ is the sample coefficient of variation (the sample coefficient of variation is well known in statistical texts; see Amiri (2016) for details). The δ_0 is constant for all the weights and by imposing the sum of weights is equal to 1, the weights under MSE is

$$\widehat{w}_{r0} = \frac{n_r \frac{\bar{X}_r}{s_r^2}}{\sum_{j=1}^{K} n_r \frac{\bar{X}_j}{s_j^2}}.$$

Unlike the GDE, it uses the sample mean in the weights. The other approach is to consider the constrained MSE:

$$MSE_\lambda(\widehat{\mu}) = (\sum_{r=1}^{K} w_r \mu_r - \mu)^2 + \sum_{r=1}^{K} w_r^2 \frac{\sigma_r^2}{n_r} - \lambda(\sum_{r=1}^{K} w_r - 1).$$

To find the minimum, take the derivative of $MSE_\lambda(\widehat{\mu})$:

$$\frac{\partial MSE_\lambda(\hat{\mu})}{\partial w_r} = 2\mu_r(\sum_{r=1}^{K} w_r \mu_r - \mu) + 2w_r \frac{\sigma_r^2}{n_r} - \lambda;$$

hence, $\frac{\partial MSE_\lambda(\widehat{\mu})}{\partial w_r} = 0$ leads to

$$w_{r0} = (\frac{\lambda}{2} - (\sum_{r=1}^{K} w_r\mu_r - \mu)\mu_r)\frac{n_r}{\sigma_r^2}.$$

If the bias is zero, $\sum_{r=1}^{K} w_r\mu_r - \mu = 0$ then $w_{r0} = (\frac{\lambda}{2})\frac{n_r}{\sigma_r^2}$. Under holding $\sum_{r=1}^{K} w_{r0} = 1$, the appropriate weights are

$$w_{r0} = \frac{\frac{n_r}{\sigma_r^2}}{\sum_{r=1}^{K}\frac{n_r}{\sigma_r^2}}.$$

The $\widehat{w}_{r0}$ is the *GDE*, which provides an unbiased estimate unlike the estimate under the unconstrained *MSE*, where $\sum_{r=1}^{K} w_r\mu_r \neq \mu$.

As mentioned earlier, the $\widehat{w}_r$s are functions of observation, so the study of underlying properties might be of interesting. We use an asymptotic technique in the statistical inference, let us define the combined *edf* as

$$\widehat{F}_{\widehat{w}}(t) = \sum_{r=1}^{K} \widehat{w}_r\widehat{F}_r(t) = \sum_{r=1}^{K} \widehat{w}_r \frac{1}{n_r}\sum_{j=1}^{n_r} I(X_{r,j} \leq t), \tag{11}$$

where $\sum \widehat{w}_r = 1$. The $\widehat{w}_r$s are a function of observations, so we present a series of asymptotic results to show their properties. If $n \to \infty$ and $\widehat{w}_r \to w_r$, $r = 1, \ldots, K$, then $\widehat{F}_{\widehat{w}}(t) \to F_w(t)$, where

$$F_w(t) = \sum_{r=1}^{K} w_r F_r(t). \tag{12}$$

To study the asymptotic analysis in depth and, possibly, obtain more relevant inference, let Γ_2 be the set of all distribution functions F with $\int x^2 dF(x) < \infty$ and define the D_2 on Γ_2 as the infimum of $\sqrt{E(|X-Y|^2)}$ over all pairs of random variables X and Y with marginal F and G, respectively. We can show the following propositions.

Proposition 2. *Suppose $F_w(.)$ is a continuous density function and $\widehat{w}_r \to w_r$. If $\widehat{F}_r(t)$ converges $F_r(t)$ for all t, almost surely (a.s.), $r = 1, \ldots, K$, then*

$$|\widehat{F}_{\widehat{w}}(t) - F_w(t)| \to 0, \text{a.s.} \forall t.$$

Proof. Assume $\widehat{F}_r(t)$ is a reasonable estimate of $F_r(t)$ which is empirical distribution function in our discussion, by applying the Glivenko–Cantelli Theorem on the empirical distribution of $\mathcal{X}_r$, can show

$$\sup_{t\in\mathbb{R}}\{\widehat{F}_r(t) - F_r(t)\} \to 0 \text{a.s.} \forall t, n_r \to \infty.$$

We can also show

$$|\widehat{F}_{\widehat{w}}(t) - F_w(t)| = |\sum_{r=1}^{k} \widehat{w}_r\widehat{F}_{(r)}(t) - \sum_{r=1}^{k} w_r F_{(r)}(t)|$$

$$\leq \sum_{r=1}^{k} |\widehat{w}_r\widehat{F}_{(r)}(t) - w_r F_{(r)}(t)|.$$

By applying Glivenko–Cantelli Theorem on the empirical distribution of $\mathcal{X}_r$, one can show following statements:

$$\sup_{t\in\mathbb{R}}\{\widehat{F}_r(t) - F_r(t)\} \to 0 \text{a.s.} \forall t, n_r \to \infty.$$

We can also show

$$|\widehat{F}_{\widehat{w}}(t) - F_w(t)| = |\sum_{r=1}^{k} \widehat{w}_r \widehat{F}_{(r)}(t) - \sum_{r=1}^{k} w_r F_{(r)}(t)|$$

$$\leq \sum_{r=1}^{k} |\widehat{w}_r \widehat{F}_{(r)}(t) - w_r F_{(r)}(t)|.$$

Because $\widehat{w}_r - w_r$ and $\widehat{w}_r\widehat{F}_r(t) - w_r F_r(t)$ converge to zero a.s. as $n_r \to \infty$, $\forall t$, therefore, we establish $|\widehat{F}_{\widehat{w}}(t) - F_w(t)| \to 0$, a.s.$\forall t$.□

Proposition 3. *If* $F_w(t) \in \Gamma_2$ *and* $\widehat{F}_{\widehat{w}}(t)$ *is defined in Equation* (11), *then the following statement is held*

$$D_2(\widehat{F}_{\widehat{w}}(t), F_w(t)) \to 0.$$

Proof. To prove, we use a kind of technique given in (Bickel and Freedman 1981); let $\widehat{F}_{\widehat{w}}, F_w \in \Gamma_2$, then $D_2(\widehat{F}_{\widehat{w}}, F_w)$ converges to zero if and only if $\widehat{F}_{\widehat{w}}$ converges to F_w in distribution, and $\int |x|^2 d\widehat{F}_{\widehat{w}}(x)$ converges to $\int |x|^2 dF_w(x)$ as n goes to ∞. The first condition is proved in Proposition 2, while the second condition follows from

$$\int t^2 d\widehat{F}_{\widehat{w}}(t) = \sum_{r=1}^{K} \widehat{w}_r \sum_{j=1}^{n_r} \frac{X_{r,j}^2}{n_r} \to \sum_{r=1}^{K} w_r \int t^2 dF_r(t)$$

$$= \int t^2 d\sum_{r=1}^{K} w_r F_r(t) = \int t^2 dF_w(t).$$

Therefore, $D_2(\widehat{F}_{\widehat{w}}, F_w) \to 0$ a.s.□

4. Statistical inference

Sections 2 and 3 are concerned with finding an accurate point estimate of the common mean from multiple samples. Here, we also discuss the confidence interval or statistical test of the suggested estimators. As mentioned, the weights are a function of observations, and developing a general classical inference for these weights is not trivial. As a result, we considered the bootstrap method, a standard tool in statistical analysis to achieve the statistical test and confidence interval. In this work, the nonparametric bootstrap is also considered that is the empirical distribution function (*edf*) that serves as a good approximation to the population distribution function. The bootstrap can be used to obtain the sampling distribution of a statistic. The bootstrap allows for estimation of the standard error of any well-defined statistic and enables us to draw inferences when the exact or the asymptotic distribution of the statistic is unavailable. The confidence interval of the proposed point estimate can be obtained as follows:
1. Generate resamples from each of the samples

$$\mathcal{X}_1^* = \{X_{1,1}^*, X_{1,2}^*, \ldots, X_{1,n_1}^*\} \overset{i.i.d.}{\sim} \widehat{F}_1(.), \tag{13}$$

$$\mathcal{X}_2^* = \{X_{2,1}^*, X_{2,2}^*, \ldots, X_{2,n_2}^*\} \overset{i.i.d.}{\sim} \widehat{F}_2(.),$$

$$\vdots$$

$$\mathcal{X}_K^* = \{X_{K,1}^*, X_{K,2}^*, \ldots, X_{K,n_K}^*\} \overset{i.i.d.}{\sim} \widehat{F}_K(.).$$

2. Calculate the estimate, $\widehat{\mu}^*$, using the proposed method on $\{\mathcal{X}_1^*, \ldots, \mathcal{X}_K^*\}$.
3. Repeat Steps 1–2 for B times and calculate $\widehat{\mu}_b^*, b = 1, \ldots, B,$ from the resamples.
4. The $V(\widehat{\mu})$ can be estimated using

$$\widehat{V}(\widehat{\mu})^* = \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{\mu}_b^* - \bar{\mu}^*)^2,$$

where $\bar{\mu}^* = \frac{1}{B}\sum_{b=1}^{B}\widehat{\mu}_b^*$.
5. Denote the ith order statistic by $\widehat{\mu}_{[i]}^*$, the confidence interval of μ at the level of $1-\alpha$ can be calculated using

$$(\widehat{\mu}^*_{[B\alpha/2]}, \widehat{\mu}^*_{[B(1-\alpha/2)]}).$$

We also explore the properties of the proposed bootstrap method.

Proposition 4. *Suppose $\widehat{F}_r$ is the edf of the rth row where the resampling plan is given in* Equaion (13) *and $F \in \Gamma_2$. Let $v_i^* =_i(\theta(\hat{F}_i^*) - \theta(\hat{F}_i))$, and then $(v_1^*, \ldots, v_K^*)$ converges in the distribution to a multivariate normal distribution $N(0, \mathrm{diag}(\sigma^2(\widehat{F}_1), \ldots, \sigma^2(\widehat{F}_K)))$ where $\sigma^2(\widehat{F}_r) = \int (X-\mu)^2 d\widehat{F}_r(X)$ and $\widehat{F}_{\widehat{w}}(x) = \sum_{r=1}^{k} \widehat{w}_r \widehat{F}_r(x)$.*

Lemma 1. *Suppose $\mathcal{X}_r^* = \{X_{r1}^*, \ldots, X_{r,n_r}^*\}$ is the bootstrap samples generated from $\widehat{F}_r(x)$ using the proposed algorithm, and let $\widehat{F}_r^*(t)$ be the edf of X_r^*. Then, $\| \widehat{F}_r^*(t) - \widehat{F}_r(t)\|_\infty = \sup_{t\in\mathbb{R}} |\widehat{F}_r^*(t) - \widehat{F}_r(t)|$ converges to zero a.s. for all $r = 1, \ldots, K,$ as n_r approaches to ∞.*

To find the relation between the distribution function of the bootstrap sample and the distribution function of population, we pursue the following result.

Proposition 5. *If $F \in \Gamma_2$ and $\widehat{F}_r^*$ is the edf defined in* (13), *then $\mathcal{D}_2(\widehat{F}^*, F)$ converges to zero a.s. as the sample size approaches to ∞.*

Proof. Lemma 1 shows $\widehat{F}_r^*(t)$ converges to $\widehat{F}_r(t)$ a.s. as $n_r \to \infty$. It follows that

$$\sup_{t\in\mathbb{R}}\{\widehat{F}_r^*(t) - F_r(t)\} = \sup_{t\in\mathbb{R}}\{\widehat{F}_r^*(t) - \widehat{F}_r(t) + \widehat{F}_r(t) - F_r(t)\}$$

$$= \sup_{t\in\mathbb{R}}\{\widehat{F}_r^*(t) - \widehat{F}_r(t)\} + \sup_{t\in\mathbb{R}}\{\widehat{F}_r(t) - F_r(t)\}$$

$$= S_1 + S_2.$$

Under Lemma 1, S_1 converges ($a.s.$) to zero, and using Glivenko–Cantelli Theorem, one can show S_2 converges ($a.s.$) to zero. Therefore, we have

$$\sup_{t\in\mathbb{R}}\{\widehat{F}_r^*(t) - F_r(t)\} \to 0 \text{as} n_r \to \infty(a.s.).$$

The second condition is satisfied because

$$\int t^2 dF^* = \sum_{r=1}^{K}\frac{1}{K}\sum_{j=1}^{n_r}\frac{x_{rj}^{*2}}{n_r} \xrightarrow{a.s} \sum_{r=1}^{K}\frac{1}{K}\sum_{j=1}^{n_r}\frac{x_{rj}^{2}}{n_r} \xrightarrow{a.s} \sum_{r=1}^{K}\frac{1}{K}\int t^2 dF_r$$

$$= \int t^2 d\sum_{r=1}^{K}\frac{1}{K}F_r = \int t^2 dF.$$

□

This proposition shows a very strong result using the bootstrap method which provides an accurate statistical result theoretically for the proposed method.

6. Numerical evaluation

In this section, we examine the performance of the proposed methods for estimating the population mean of independent samples, but not identical. To evaluate our proposed methodologies, we carried out three numerical analyses: one with simulated data from a normal distribution and the others using real data where they are generated from populations with different distributions. Technically, we know their true mean, and it can be used to evaluate the discussed methods using these data sets. Thus, the question is: how well the proposed methods can estimate the mean of the underlying population to compare the estimates, we consider the mean square error (MSE): we generate $M = 10{,}000$ random samples for the given Ks corresponding to the designs and estimate the means via the discussed methods. Let us denote the estimate of mean as $\widehat{\mu}_m, m = 1, \ldots, M$ which can be computed using the discussed methods, since the μs of the underlying populations are known, the estimate $\widehat{MSE}$ for the given μ is computed using $\widehat{MSE}(\widehat{\mu}) = \frac{1}{M}\sum_{m=1}^{M}(\widehat{\mu}_m - \mu)^2$. We consider the GDE, the standard method, and the other competing estimates are compared with it using the relative accuracy (RA),

$$RA(\widehat{\mu}^*, \widehat{\mu}_{GDE}) = \frac{\widehat{MSE}(\widehat{\mu}^*)}{\widehat{MSE}(\widehat{\mu}_{GDE})}. \tag{14}$$

$RA(\widehat{\mu}^*, \widehat{\mu}_{GDE}) < 1$ implies the $\widehat{\mu}^*$ outperforms the $\widehat{\mu}_{GDE}$.

6.1. Normality assumption

In this section, we study the proposed methods under holding the normality of distribution. To achieve this aim, they are compared under the normal distributions with different parameters; we set $\mu = 2$ and different values for $\sigma_r, r = 1, \ldots, K$ are used. We generate random samples for the given $K = 2$, 5, and 10. For $K = 2$, we consider $(n_1, n_2) = (8, 16)$, and the designs are as follows:

$$\mathcal{D}_{N1} = \{(\sigma_1^2, \sigma_2^2) = (1, 1)\},$$

$$\mathcal{D}_{N2} = \{(\sigma_1^2, \sigma_2^2) = (3, 3)\},$$

$$\mathcal{D}_{N3} = \{(\sigma_1^2, \sigma_2^2) = (1, 3)\},$$

$$\mathcal{D}_{N4} = \{(\sigma_1^2, \sigma_2^2) = (3, 1)\}.$$

For $K = 5$, we set $(n_1, n_2, n_3, n_4, n_5) = (8, 10, 13, 16, 18)$ with the following designs:

$$\mathcal{D}_{N5} = \{(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2) = (1, 1, 1, 1, 1)\},$$

$$\mathcal{D}_{N6} = \{(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2) = (3, 3, 3, 3, 3)\},$$

$$\mathcal{D}_{N7} = \{(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2) = (1, 1, 1, 3, 3)\},$$

$$\mathcal{D}_{N8} = \{(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2) = (3, 3, 1, 1, 1)\},$$

$$\mathcal{D}_{N9} = \{(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2) = (1, 1.5, 2, 2.5, 3)\},$$

$$\mathcal{D}_{N10} = \{(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2) = (1, 2.5, 2, 1.5, 3)\},$$

$$\mathcal{D}_{N11} = \{(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2) = (3, 1.5, 2, 2.5, 1)\},$$

$$\mathcal{D}_{N12} = \{(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2) = (2.5, 1.5, 2, 1, 3)\},$$

$$\mathcal{D}_{N13} = \{(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2) = (3, 1, 2, 2.5, 1.5)\},$$

$$\mathcal{D}_{N14} = \{(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2) = (1.5, 2, 3, 2.5, 1)\}.$$

For $K = 10$, we set $(n_1, n_2, \ldots, n_{10}) = (8, 10, 13, 16, 18, 20, 23, 26, 29, 33)$ with the following designs:

$$\mathcal{D}_{N15} = \{(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{10}^2) = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)\},$$

$$\mathcal{D}_{N16} = \{(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{10}^2) = (1, 1, 1, 1, 1, 3, 3, 3, 3, 3)\},$$

$$\mathcal{D}_{N17} = \{(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{10}^2) = (3, 3, 3, 3, 3, 1, 1, 1, 1, 1)\},$$

$$\mathcal{D}_{N18} = \{(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{10}^2) = (1, 3, 1, 3, 1, 3, 1, 3, 1, 3)\}.$$

The *RA* under the discussed designs are given in Table 1. Obviously under $\mathcal{D}_{N1}$, $\mathcal{D}_{N2}$, $\mathcal{D}_{N5}$, and $\mathcal{D}_{N6}$, where the samples are generated from the same parent populations, *PM* has the highest efficiency as shown in the discussion in Section 2. However, under the other designs, where samples generated from the nonidentical distributions, *DGDE* provides a good estimator. Interestingly, *DSV* has a good performance under $\mathcal{D}_{N5}$ and $\mathcal{D}_{N6}$. So the *DGDE* can be a convenient method for estimating the mean; however, the MLE does not perform better than the *DGDE* for the proposed designs, and their performances are close. The next subsection considers skewed real data; it is shown that *DSV* might provide a reasonable estimate.

The new estimate is a weighted mean in terms of *DGDE* and *DSV*, a function of observations; hence, we use the bootstrap method to estimate the variance of *DGDE* ($\widehat{V}^*(\widehat{\mu}_{DGDE})$) and the variance *SV* ($\widehat{V}^*(\widehat{\mu}_{DSV})$), and then select the weighted estimator:

$$\widehat{\mu}_{DGSV} = W_0 \widehat{\mu}_{DGDE} + (1 - W_0) \widehat{\mu}_{DSV},$$

where

Table 1. The performance of proposed methods on simulated data from the normal distributions.

	$K=2$						
	CM	*PM*	MLE	*SV*	*DGDE*	*DSV*	*DGSV*
$\mathcal{D}_{N1}$	0.96	0.91	1.02	1.00	1.01	1.01	1.01
$\mathcal{D}_{N2}$	0.96	0.91	1.02	1.00	1.01	1.01	1.01
$\mathcal{D}_{N3}$	1.81	2.51	1.01	1.00	1.00	1.00	1.00
$\mathcal{D}_{N4}$	3.99	2.54	0.97	1.00	0.98	0.98	0.98
	$K=5$						
	CM	*PM*	MLE	*SV*	*DGDE*	*DSV*	*DGSV*
$\mathcal{D}_{N5}$	0.93	0.85	0.98	0.90	0.98	0.89	0.90
$\mathcal{D}_{N6}$	0.92	0.83	0.97	0.89	0.98	0.88	0.90
$\mathcal{D}_{N7}$	1.64	2.40	0.99	1.24	1.00	1.24	1.00
$\mathcal{D}_{N8}$	3.95	2.18	0.99	2.30	0.99	2.11	1.12
$\mathcal{D}_{N9}$	1.04	1.36	1.01	0.98	1.00	0.98	0.94
$\mathcal{D}_{N10}$	1.27	1.42	1.00	1.11	1.00	1.09	0.97
$\mathcal{D}_{N11}$	2.21	1.64	0.97	1.63	0.98	1.56	1.07
$\mathcal{D}_{N12}$	1.74	1.67	0.97	1.47	0.97	1.43	1.04
$\mathcal{D}_{N13}$	1.80	1.42	0.99	1.29	0.99	1.23	1.01
$\mathcal{D}_{N14}$	1.78	1.69	0.97	1.57	0.97	1.55	1.06
	$K=10$						
	CM	*PM*	MLE	*SV*	*DGDE*	*DSV*	*DGSV*
$\mathcal{D}_{N15}$	1.06	0.87	0.96	1.01	0.97	1.00	0.93
$\mathcal{D}_{N16}$	1.48	2.19	0.97	1.35	0.97	1.35	1.00
$\mathcal{D}_{N17}$	5.06	2.36	0.99	4.11	0.99	3.95	1.14
$\mathcal{D}_{N18}$	2.58	2.43	0.96	2.16	0.96	2.11	1.08

$$W_0 = \frac{\widehat{V}^*(\widehat{\mu}_{DSV})}{\widehat{V}^*(\widehat{\mu}_{DSV}) + \widehat{V}^*(\widehat{\mu}_{DGDE})}.$$

Essentially, the $\widehat{\mu}_{DGSV}$ is a weighted combination of *DGDE* and *DSV*. Table 1 shows that the $\widehat{\mu}_{DGSV}$ performs well and can be nominated as a good estimator for the common mean. A benefit of this technique is that we do not need to find a formal condition where either *DGDE* or *DSV* works accurately.

6.2. Random effects

In this section, the performance of the proposed method under holding the random-effect is studied, unlike the fixed effect model, the parameters of model follow some distributions. Such model is considered in the medical sciences (see Higgins et al. (2009), Bodnar et al. (2017), and Guolo and Varin (2017) as well as references therein. The random effects model is defined as

$$X_{r,j} = \mu + \lambda_r + \epsilon_{rj} \tag{15}$$

$$\lambda_r \sim N(0, \sigma_\tau^2),$$

$$\epsilon_{rj} \sim N(0, \sigma^2)$$

$$\sigma_\tau^2 \sim \mathrm{Unif}(0, \tau).$$

To study the performance, we consider the the following designs for $K = 5$ with $(n_1, n_2, n_3, n_4, n_5) = (8, 10, 13, 16, 18)$:

$$\mathcal{D}_{R1} = (\mu = 2, \tau = 1, \sigma = 1), \tag{16}$$

$$\mathcal{D}_{R2} = (\mu = 2, \tau = 1, \sigma = 2),$$

$$\mathcal{D}_{R3} = (\mu = 2, \tau = 2, \sigma = 1),$$

$$\mathcal{D}_{R4} = (\mu = 2, \tau = 2, \sigma = 2),$$

$$\mathcal{D}_{R5} = (\mu = 2, \tau = 2, \sigma = 3),$$

$$\mathcal{D}_{R6} = (\mu = 2, \tau = 3, \sigma = 2),$$

$$\mathcal{D}_{R7} = (\mu = 2, \tau = 3, \sigma = 3),$$

The result of running the proposed methods on the data is presented in Table 2, it shows *DSV* and *DGSV* have the best performance among the proposed methods.

6.3. Diabetes

Next, we turn to the comparisons of the methods on clinical data. The data are Pima Indians Diabetes Data Set, which can be found on the UC Irvine Machine Learning Repository,[1] and the diabetes is a metabolic disease, which make cause high glucose over a prolonged period. It is a growing disease because of lack of any severe symptom at early stages and it is also called a silent killer. This data set includes female patients at least 21 years old of Pima Indian heritage. The summary statistics, min, first quartile, median, third quartile, and max are 0.0780, 0.2438, 0.3725, 0.4719, 0.6262, and 2.4200, respectively.

To explore the performance of proposed data, we consider *Diabetes pedigree function*, which is very skewed data. We consider five subsets of this data set: *PI* (data between 25% and 75% percentiles), *PII* (data between 20% and 80% percentiles), *PIII* (data between 15% and 85% percentiles), *PIV* (data between 10% and 90% percentiles), and *PV* (data between 5% and 95% percentiles). To have data sets with the same mean, the values are divided by its mean and multiplied by the mean of the original data, this design helps to mimic the complex structure of data. Their standard deviations are 0.1355, 0.1591, 0.1846, 0.2129 and 0.2494. Figure 1 displays the histograms of data and generated data. Unlike the simulated data in the previous subsection from the normal distributions, this data setup helps to evaluate the proposed methods under data generated from different distributions.

Table 2. The performance of proposed methods on the random effects model.

Design	*CM*	*PM*	MLE	*SV*	*DGDE*	*DSV*	*DGSV*
$\mathcal{D}_{R1}$	0.84	0.88	0.89	0.84	0.89	0.79	0.79
$\mathcal{D}_{R2}$	0.86	0.86	0.96	0.86	0.96	0.84	0.86
$\mathcal{D}_{R3}$	0.83	0.88	0.85	0.84	0.81	0.72	0.72
$\mathcal{D}_{R4}$	0.82	0.87	0.91	0.83	0.90	0.78	0.80
$\mathcal{D}_{R5}$	0.86	0.88	0.94	0.86	0.94	0.83	0.85
$\mathcal{D}_{R6}$	0.82	0.87	0.83	0.84	0.84	0.75	0.75
$\mathcal{D}_{R7}$	0.83	0.87	0.88	0.84	0.89	0.78	0.79

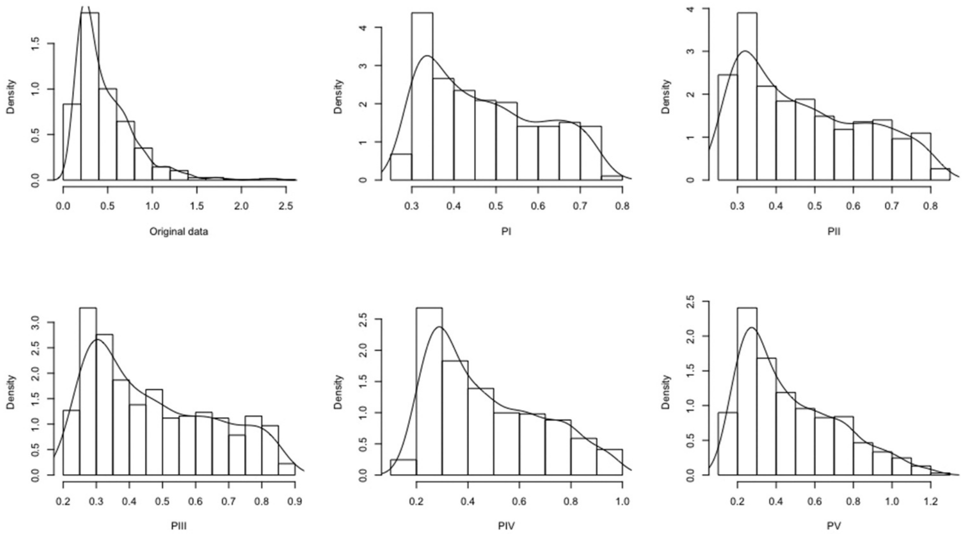

Figure 1. The histograms of Diabetes pedigree function and generated data sets; $PI, \ldots, PV$.

Table 3. The performance of proposed methods on the Diabetes pedigree function.

Design	*CM*	*PM*	MLE	*SV*	*DGDE*	*DSV*	*DGSV*
$\mathcal{D}_{D1}$	0.71	0.65	0.91	0.73	0.94	0.72	0.77
$\mathcal{D}_{D2}$	0.61	0.56	0.89	0.63	0.94	0.62	0.70
$\mathcal{D}_{D3}$	0.74	0.91	0.95	0.73	0.97	0.73	0.79
$\mathcal{D}_{D4}$	1.28	0.85	0.84	1.04	0.87	0.98	0.86
$\mathcal{D}_{D5}$	0.63	0.74	0.96	0.66	0.98	0.66	0.75
$\mathcal{D}_{D6}$	0.72	0.76	0.94	0.73	0.96	0.72	0.78
$\mathcal{D}_{D7}$	0.90	0.71	0.85	0.82	0.89	0.78	0.78
$\mathcal{D}_{D8}$	0.82	0.78	0.88	0.79	0.92	0.77	0.79
$\mathcal{D}_{D9}$	0.86	0.71	0.88	0.77	0.91	0.74	0.77
$\mathcal{D}_{D10}$	0.80	0.75	0.88	0.80	0.94	0.79	0.80

We consider the $PI, \ldots, PV$ as populations and generate 10,000 data sets using the designs in $\mathcal{D}_{D1}, \ldots, \mathcal{D}_{D10}$, (17), and then estimate the mean using the proposed method on these data sets, the result of our analyses are presented in Table 3, clearly, the best methods were PM, SV, DSV, and $DGSV$, which to perform well on this data. The key inference from our numerical work is that among the methods tested on our examples, $DGSV$ is a good option for the skew data.

$$\mathcal{D}_{D1} = \{(\mathcal{X}_1, \ldots, \mathcal{X}_5) = (PI, PI, PI, PI, PI)\}, \tag{17}$$

$$\mathcal{D}_{D2} = \{(\mathcal{X}_1, \ldots, \mathcal{X}_5) = (PV, PV, PV, PV, PV)\},$$

$$\mathcal{D}_{D3} = \{(\mathcal{X}_1, \ldots, \mathcal{X}_5) = (PI, PI, PI, PV, PV)\},$$

$$\mathcal{D}_{D4} = \{(\mathcal{X}_1, \ldots, \mathcal{X}_5) = (PV, PV, PI, PI, PI)\},$$

$$\mathcal{D}_{D5} = \{(\mathcal{X}_1, \ldots, \mathcal{X}_5) = (PI, PII, PIII, PIV, PV)\},$$

$$\mathcal{D}_{D6} = \{(\mathcal{X}_1, \ldots, \mathcal{X}_5) = (PI, PIV, PIII, PII, PV)\},$$

$$\mathcal{D}_{D7} = \{(\mathcal{X}_1, \ldots, \mathcal{X}_5) = (PV, PII, PIII, PIV, PI)\},$$

$$\mathcal{D}_{D8} = \{(\mathcal{X}_1, \ldots, \mathcal{X}_5) = (PIV, PII, PIII, PI, PV)\}.$$

$$\mathcal{D}_{D9} = \{(\mathcal{X}_1, \ldots, \mathcal{X}_5) = (PV, PI, PIII, PIV, PII)\}.$$

$$\mathcal{D}_{D10} = \{(\mathcal{X}_1, \ldots, \mathcal{X}_5) = (PII, PIII, PV, PIV, PI)\}.$$

7. Conclusion

A considerable amount of research has been done to elaborate on the statistical inference of common means when data are collected from different studies. Because researchers are able to combine different studies, more samples are available, which helps us draw a better inference. This paper has attempted to explore different approaches to estimate the common mean and compare them via the *MSE* of estimators. Unlike existing research, we consider a general setup not restricted to the parametric approach, which is of the utmost importance when working with real data.

To evaluate the proposed methods' performance, we examined them on both real and simulated data that support the intuition behind the technique. Under the normality of data, *MLE* and *DGDE* performances are very close to *GDE*, but *DGSV* outperforms the others. Section 5.2 studied the proposed method on the random effects model, and it shows *DGSV* got a more accurate estimate than others. Based on the numerical exploration in Section 5.3, where the real data is skewed, *DGSV* outperforms *MLE*. Our findings indicate that *DGSV* performs well compared to the other methods. This nonparametric estimator does not require holding the normality of populations.

In addition to proposing new methods for estimating the common mean, we also present statistical inferences of proposed methods that provide formal justification to show why they should work well in some generality. What we deliver is a nonparametric and wholly computational resampling method. As dedicated supporters of open-science, we have deployed a git[2] to release, maintain, and support the complete end-to-end implementation of all steps of the proposed estimates.

Notes

1. https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes
2. https://saeidamiri1.github.io/codes/mtmeans/mtmeans

References

Amiri, S. 2016. Revisiting inference of coefficient of variation: Nuisances parameters. *Stat* 5 (1):234–241. doi:10.1002/sta4.116.

Bickel, P. J., and D. A. Freedman. 1981. Some asymptotic theory for the bootstrap. In *The annals of statistics*, 1196–1217.

Bodnar, O., A. Link, B. Arendacká, A. Possolo, and C. Elster. 2017. Bayesian estimation in random effects meta-analysis using a non-informative prior. *Statistics in Medicine* 36 (2):378–399. doi:10.1002/sim.7156.

Brockwell, S. E., and I. R. Gordon. 2001. A comparison of statistical methods for meta-analysis. *Statistics in Medicine* 20 (6):825–840. doi:10.1002/sim.650.

Chang, C.-H., and N. Pal. 2008. Testing on the common mean of several normal distributions. *Computational Statistics & Data Analysis* 53 (2):321–333. doi:10.1016/j.csda.2008.07.024.

Chang, Y.-T., Y. Oono, and N. Shinozaki. 2012. Improved estimators for the common mean and ordered means of two normal distributions with ordered variances. *Journal of Statistical Planning and Inference* 142 (9):2619–2628. doi:10.1016/j.jspi.2012.03.006.

Graybill, F. A., and R. Deal. 1959. Combining unbiased estimators. *Biometrics* 15 (4):543–550. doi:10.2307/2527652.

Guolo, A., and C. Varin. 2017. Random-effects meta-analysis: The number of studies matters. *Statistical Methods in Medical Research* 26 (3):1500–1518. doi:10.1177/0962280215583568.

Hartung, J., G. Knapp, and B. K. Sinha. 2011. *Statistical meta-analysis with applications*, Vol. 738. John Wiley & Sons.

Higgins, J. P. T., S. G. Thompson, and D. J. Spiegelhalter. 2009. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172 (1):137–159. doi:10.1111/j.1467-985X.2008.00552.x.

Olkin, I., and Keller, T. 2004. Combining correlated unbiased estimators of the mean of a normal distribution. In A festschrift for Herman Rubin. Institute of Mathematical Statistics, 218-227..

Kim, M. G. 2013. Influence analysis of the common mean problem. *Communications for Statistical Applications and Methods* 20 (3):217–223. doi:10.5351/CSAM.2013.20.3.217.

Ma, T., R. Ye, and L. Jia. 2011. Finite-sample properties of the Graybill–Deal estimator. *Journal of Statistical Planning and Inference* 141 (11):3675–3680. doi:10.1016/j.jspi.2011.06.004.