



# Uniformly consistently estimating the proportion of false null hypotheses via Lebesgue–Stieltjes integral equations

Xiongzhi Chen

Department of Mathematics and Statistics, Washington State University, Pullman, WA 99164, USA

## ARTICLE INFO

### Article history:

Received 10 September 2018  
Received in revised form 17 June 2019  
Accepted 17 June 2019  
Available online 26 June 2019

### AMS 2010 subject classifications:

primary 35C05  
60E05  
secondary 60F10

### Keywords:

Analytic functions  
Bessel functions  
Concentration inequalities  
Fourier transform  
Lambert W functions  
Lebesgue–Stieltjes integral equations  
Mellin transform  
Natural exponential family  
Proportion of false null hypotheses

## ABSTRACT

The proportion of false null hypotheses is a very important quantity in statistical modelling and inference based on the two-component mixture model and its extensions, and in control and estimation of the false discovery rate and false non-discovery rate. Most existing estimators of this proportion threshold  $p$ -values, deconvolve the mixture model under constraints on its components, or depend heavily on the location-shift property of distributions. Hence, they usually are not consistent, applicable to non-location-shift distributions, or applicable to discrete statistics or  $p$ -values. To eliminate these shortcomings, we construct uniformly consistent estimators of the proportion as solutions to Lebesgue–Stieltjes integral equations. In particular, we provide such estimators respectively for random variables whose distributions have Riemann–Lebesgue type characteristic functions, whose distributions form discrete natural exponential families with infinite supports, and whose distributions form natural exponential families with separable moment sequences. We provide the speed of convergence and uniform consistency class for each such estimator under independence. In addition, we provide two examples for which a consistent estimator of the proportion cannot be constructed using our techniques.

© 2019 Elsevier Inc. All rights reserved.

## 1. Introduction

The proportion of false null hypotheses and its dual, the proportion of true null hypotheses, play important roles in statistical modelling and multiple hypotheses testing. For example, they are components of the two-component mixture model of [14], its extensions by [5,28] and [35], and their induced statistics including the “local false discovery rate (local FDR)” of [14] and “positive FDR (pFDR)” and  $q$ -value of [39]. They also form the optimal discovery procedure (ODP) of [40]. However, without information on the proportions, decision rules based on the local FDR cannot be implemented, and none of the pFDR,  $q$ -value and ODP decision rule can be computed in practice. On the other hand, these proportions form upper bounds on the FDRs and false non-discovery rates (FNRs) of all FDR procedures including those of [2,3,17,37] and [41]. Therefore, accurate information on either proportion helps better control and estimate the FDR and FNR, and thus potentially enables an adaptive FDR procedure to be more powerful or have smaller FNR than its non-adaptive counterpart. Since neither proportion is known in practice, it is very important to accurately estimate the proportions.

In this work, we focus on consistently estimating the proportions without employing the two-component mixture model, requiring  $p$ -values to be identically distributed under the alternative hypotheses, or assuming that statistics or  $p$ -values have absolutely continuous cumulative distribution functions (CDFs). The main motivation for dealing with

E-mail address: [xiongzhi.chen@wsu.edu](mailto:xiongzhi.chen@wsu.edu).

discrete p-values or statistics is the wide practise of FDR control in multiple testing based on discrete data in genomics [1,36], genetics [18], vaccine efficacy studies [31], drug safety monitoring [8] and other areas, where Binomial test (BT), Fisher's exact test (FET) and exact Negative Binomial test (NBT) have been routinely used to test individual hypotheses. In the sequel, we refer to an estimator of the proportion of true (or false) null hypotheses as a “null (or alternative) proportion estimator”.

### 1.1. A brief review on existing proportion estimators

There are many estimators of the proportions, and their constructions can be roughly categorized into 4 classes: (i) thresholding p-values [38,41]; (ii) deconvolving the two-component mixture model for the distribution of p-values or statistics, modulo identifiability conditions [17,24,42]; (iii) bounding the proportions via the use of uniform empirical process [32]; (iv) Fourier transform for Gaussian family or mixtures with a Gaussian component [20–22]. Further, perhaps the only existing consistent proportion estimators are those mentioned right above, and only those in class (iii) and (iv) are consistent when the proportion tends to zero (see (3) for the definition of consistency). However, these estimators have three disadvantages: the construction in class (iv) based on Fourier transform does not work if the family of distributions of statistics does not consist at least one component from a location-shift family; consistency of estimators in class (i) and (ii) requires the two-component mixture model and various regularity conditions such as concavity, smoothness, purity (defined by [17]) of the alternative component, or the conditions in Lemma 3 or Lemma 4 of [24]; consistency of the estimator in class (iii) requires p-values to have continuous distributions, be independent, and be identically distributed under the alternative hypotheses. An extended comparison for popular proportion estimators is provided in Table 1.

In practise, the requirements needed by the consistent proportion estimators mentioned above are often not met. For example, for multiple testing where each statistic follows a Chi-square or Binomial distribution the involved distributions do not form a location-shift family, the two-component mixture model is inappropriate for p-values of BTs, FETs or NBTs with different marginal counts, and p-values under the alternative hypotheses cannot be identically distributed when the signal levels are different across individual hypotheses under the alternative. Further, almost all existing proportion estimators were initially designed for p-values or statistics that have continuous distributions, and it is not clear yet whether the null proportion estimators in classes (ii) to (iv) are reciprocally conservative even though the estimator of [41] is so for independent p-values whose null distributions are uniform on  $[0, 1]$  (see Corollary 13 of [3]). Therefore, it is important to develop new, consistent proportion estimators that mitigate or eliminate the shortcomings of existing ones.

### 1.2. Main contributions and summary of results

We generalize “Jin’s strategy” provided in Section 2.1 of [20] that estimates the proportions by approximating the indicator function of the true status of a null hypothesis and has only been implemented for Gaussian family or mixtures with a Gaussian component, and in much more general settings construct proportion estimators as solutions to a specific type of Lebesgue–Stieltjes integral equation in the complex domain. Generalizing Jin’s strategy to non-location-shift families is highly nontrivial since this requires solving Lebesgue–Stieltjes integral equations and outside location-shift families solutions to these equations are very hard to find or do not exist. Even after a solution is found, to prove the uniform consistency of the proposed estimator we have to derive concentration inequalities for non-Lipschitz transforms of independent, unbounded random variables, and this is highly nontrivial and an open area in probability theory.

Our methodology (referred to as “the Strategy”) lifts Jin’s strategy to its full generality. In addition to the key advantage of Jin’s strategy, i.e., being independent from but partially applicable to the two-component mixture model, the Strategy applies to random variables that have continuous or discrete distributions or form non-location-shift families. Further, it produces uniformly consistent proportion estimators in the dense and moderately sparse regimes (that are defined at the beginning of Section 2).

Our main contributions are summarized as follows:

- “Construction I”: Construction of uniformly consistent proportion estimators for random variables whose distributions have Riemann–Lebesgue type characteristic functions (RL type CFs); see Definition 1 in Section 3. In particular, this covers the construction for random variables from several location-shift families, including that in [20] for Gaussian and Laplace distributions.
- “Construction II”: Construction of uniformly consistent proportion estimators for random variables whose distributions form discrete natural exponential families (NEFs) with infinite supports; see Sections 4 and 6. In particular, this covers six of the twelve natural exponential families with cubic variance functions (NEF-CVFs) proposed by [27] such as Poisson and Negative Binomial families. Estimators from Construction II are perhaps the only consistent estimators that are applicable to discrete random variables.
- “Construction III”: Construction of uniformly consistent proportion estimators for random variables whose distributions form continuous NEFs with separable moment sequences; see Sections 4 and 7. In particular, this covers Gamma family which includes Exponential and central Chi-square families as special cases.

**Table 1**

A comparison of some popular estimators of the proportion of false null hypotheses. The 2nd column of the table provides features of the estimators stated in its corresponding 1st column, and the 3rd column of the table details on items in the 2nd column.

Thresholding estimators	References	[9,38,41]
	Method	Thresholding $p$ -values
	Disadvantages	Only applicable to $p$ -values
	Consistency	Consistency when $p$ -values are independent, identically distributed, follow an identifiable, two-component mixture model, and have an absolutely continuous CDF with some regularity properties, which has been proved by [17] for Storey's estimator
	Scope	Applicable to both discrete and continuous data
Deconvolution estimators: I	References	[17,42]
	Method	Deconvolving $p$ -value densities
	Modelling assumptions	$p$ -values need to be identically distributed, follow a two-component mixture model, and have an absolutely continuous CDF
	Consistency	Joint assumptions for consistency: independence, an identifiable model, regularity conditions on the alternative component in the model, and the proportion being fixed
	Scope	Applicable to continuous data only
Deconvolution estimators: II	References	[24]
	Method	Deconvolving $p$ -value densities
	Modelling assumptions	$p$ -values need to be identically distributed and follow a two-component mixture model
	Consistency	Joint assumptions for consistency: independence, an identifiable model, regularity conditions on the alternative component in the model, and the proportion being fixed
	Scope	Applicable to both discrete and continuous data
Deconvolution estimators: III	References	[25]
	Method	Deconvolving $p$ -value densities
	Modelling assumptions	$p$ -values need to be identically distributed, follow a two-component mixture model, and have an absolutely continuous, concave CDF
	Consistency	Consistency unknown
	Scope	Applicable to continuous data only
Fourier transform based estimators	References	[7,20–22]
	Method	Fourier transform and convolution
	Modelling assumptions; disadvantages	The family of distributions of statistics should consist at least one component from a location-shift family; implementation needs information on the scale parameter of the location-shift component
	Consistency	Uniform consistency under weak dependence when the proportion is fixed or converges to 0 at certain rate
	Scope	Applicable to continuous data only
Goodness-of-fit estimators	References	[32]
	Method	Measuring the relative excess of the uniform empirical process
	Modelling assumptions	$p$ -values need to have absolutely continuous CDFs, be independent, and be identically distributed under the alternative hypotheses
	Consistency	Consistency under independence when the proportion is fixed or converges to 0 at certain rate
	Scope	Applicable to continuous data only
Integro-equation based estimators	References	This paper
	Method	Approximating the status of each hypothesis via solutions to Lebesgue–Stieltjes integral equations
	Modelling assumptions and disadvantages	Statistics need to have CDFs that have Riemann–Lebesgue type characteristic functions, that are members of an NEF with support $\mathbb{N}$ , or that are members of a natural exponential family (NEF) with separable moment sequence; implementation needs information on the scale parameters of the CDFs, on the supremum norm of the parameter vector, or on the infimum of a transform of this vector
	Consistency	Uniform consistency under independence when the proportion is fixed or converges to 0 at certain rate
	Scope	Applicable to both continuous and discrete data

- For each constructed estimator, its speed of convergence and uniform consistency class is provided under independence; see Sections 3, 6 and 7.
- Concentration inequalities for sums of certain non-Lipschitz transforms of independent but not necessarily almost surely bounded random variables.

Specifically, Construction I employs Fourier transform and an extended Riemann–Lebesgue lemma of [12]. Since the set of distributions with RL type CFs contains several location-shift families, we give a unified treatment of proportion estimators for location-shift families and reveal the intrinsic mechanism of the Fourier transform based construction. In contrast, Construction II mainly uses generating functions (GFs), and Construction III Mellin transform which can be regarded as inducing “multiplication-convolution equivalence”, in contrast to Fourier transform inducing “translation-convolution equivalence”; see Section 4.5 for details on this. As negative results, we show that the Strategy is not implementable for Inverse Gaussian family (whose densities have a very special structure; see [27] for a definition of the family) and Binomial family (which is discrete but with a finite support); see Section 5.

We provide upper bounds on the variances of the proportion estimators, show their uniform consistency, and provide their speeds of convergence for consistency under independence. Additionally, we have found the following. For estimators given by Construction I, uniform consistency in frequency domain (see Definition 3) can also be achieved due to the global Lipschitz property of the construction, and their uniform consistency classes (see Definition 3) can be ordered via set inclusion according to the magnitudes of the moduli of the corresponding CFs; see Theorem 3 and Corollary 4. In contrast, for estimators given by Constructions II and III, uniform consistency in frequency domain is very hard to obtain since the constructions are not Lipschitz transforms of the involved random variables.

For an estimator given by Construction I or Construction II where GFs have finite radii of convergence, its speed of convergence and uniform consistency class do not depend on the supremum norm of the parameter vector (see Theorem 3, Corollaries 4 and 5), enabling the estimator to be fully data-adaptive. In contrast, for estimators given by other instances of Construction II and by Construction III, their speeds of convergence and uniform consistency classes may depend explicitly on the supremum norm of the parameter vector or the infimum of a transform of this vector (see Theorems 7 and 9). However, since the Strategy approximates indicator functions of the status of individual null hypotheses, these speeds and classes always depend on the minimal magnitude of the differences between parameters of interest and their reference value regardless of which among the three constructions is used – a universality phenomenon for the Strategy.

### 1.3. Notations and conventions

Throughout the article, we use the following conventions and notations:  $C$  denotes a generic, positive constant whose values may differ at different occurrences;  $O(\cdot)$  and  $o(\cdot)$  are respectively Landau's big  $O$  and small  $o$  notations;  $\mathbb{E}$  and  $\mathbb{V}$  are respectively the expectation and variance with respect to the probability measure  $\Pr$ ;  $\mathbb{R}$  and  $\mathbb{C}$  are respectively the set of real and complex numbers;  $\Re$ ,  $\Im$  and  $\arg$  denote the real, imaginary parts and argument of a complex number, respectively;  $\mathbb{N}$  denotes the set of non-negative integers, and  $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$ ;  $\delta_y$  is the Dirac mass at  $y \in \mathbb{R}$ ;  $\nu$  the Lebesgue measure, and when it is clear that an integral is with respect to  $\nu$ , the usual notation  $d\cdot$  for differential will be used in place of  $\nu(d\cdot)$ ; for  $1 \leq p \leq \infty$  and  $A \subseteq \mathbb{R}$ ,  $\|f\|_p = \left\{ \int_A |f(x)|^p \nu(dx) \right\}^{1/p}$  and  $L^p(A) = \{f : \|f\|_p < \infty\}$ , for which  $\|f\|_\infty$  is the essential supremum of  $f$ ; for a set  $A$  in  $\mathbb{R}$  and a scalar  $a$ ,  $1_A$  is the indicator of  $A$  and  $A - a = \{x - a : x \in A\}$ ;  $\emptyset$  is the empty set; for  $x \in \mathbb{R}$ ,  $[x]$  denotes the integer part of  $x$ ; the extended complex plane  $\mathbb{C} \cup \{\infty\}$  is identified with the Riemann sphere so that  $\infty$  corresponds to its north pole; for  $x \in \mathbb{R}$ ,  $x \rightarrow \infty$  means  $x \rightarrow +\infty$ , and  $x \rightarrow -\infty$  means  $x \rightarrow -\infty$ ; for positive functions  $a'(x)$  and  $b'(x)$ ,  $a'(x) \sim b'(x)$  means that  $\lim_{x \rightarrow \infty} a'(x) \{b'(x)\}^{-1} = 1$ ;  $\mathbb{R}^{\aleph}$  is the  $\aleph$ -Cartesian product of  $\mathbb{R}$ , where  $\aleph$  is the cardinality of  $\mathbb{N}$ .

### 1.4. Organization of paper

The rest of the article is organized as follows. In Section 2 we formulate the problem of proportion estimation and state the Strategy for constructing proportion estimators. In Section 3 we develop uniformly consistent proportion estimators when the CDFs of random variables have Riemann–Lebesgue type characteristic functions. In Section 4, we construct proportion estimators when the distributions of random variables form discrete NEFs with infinite supports or form NEFs with separable moment sequences. In Section 5, we provide two families of random variables for which the Strategy cannot be applied. In Sections 6 and 7 we justify the uniform consistency of the constructed proportion estimators for random variables whose distributions form NEFs with infinite supports or with separable moment sequences, respectively. In Section 8 we conduct a simulation study on the proposed estimators, with comparison to those of [32] and [20]. We end the article with a discussion in Section 9, together with several topics worthy of future investigations. The supplementary material contains proofs of all theoretical results, provides evidence that Construction III does not apply to Ressel or Hyperbolic Cosine families (defined by [27]), and a discussion on uniform consistency of proportion estimators in frequency domain for Construction II and III.

## 2. The estimation problem and strategy

In this section, we formulate the estimation problem and generalize Jin's strategy. For each  $i \in \{1, \dots, m\}$ , let  $z_i$  be a random variable with mean or median  $\mu_i$ , such that, for a fixed value  $\mu_0$  for the mean or median and some integer  $m_0$  between 0 and  $m$ ,  $\mu_i = \mu_0$  for each  $i \in \{1, \dots, m_0\}$  and  $\mu_i \neq \mu_0$  for each  $i \in \{m_0 + 1, \dots, m\}$ . Consider simultaneously testing the null hypothesis  $H_{i0} : \mu_i = \mu_0$  versus the alternative hypothesis  $H_{i1} : \mu_i \neq \mu_0$  for all  $i \in \{1, \dots, m\}$ . Let  $I_{0,m} = \{i \in \{1, \dots, m\} : \mu_i = \mu_0\}$  and  $I_{1,m} = \{i \in \{1, \dots, m\} : \mu_i \neq \mu_0\}$ . Then the cardinality of  $I_{0,m}$  is  $m_0$ , the proportion of

true null hypothesis (“null proportion” for short) is defined as  $\pi_{0,m} = m^{-1}m_0$ , and the proportion of false null hypotheses (“alternative proportion” for short)  $\pi_{1,m} = 1 - \pi_{0,m}$ . In other words,  $\pi_{0,m}$  is the proportion of random variables that have a prespecified mean or median. Our target is to consistently estimate  $\pi_{1,m}$  as  $m \rightarrow \infty$  when  $\{z_i\}_{i=1}^m$  are independent.

We will adopt the following convention from [20]: the dense regime is represented by  $\pi_{1,m}$  such that  $\liminf_{m \rightarrow \infty} \pi_{1,m} > 0$ , the moderately sparse regime by  $\pi_{1,m} = Cm^{-s}$  for  $s \in (0, 0.5)$ , the critically sparse regime by  $\pi_{1,m} = Cm^{-0.5}$ , and the very sparse regime by  $\pi_{1,m} = Cm^{-s}$  for  $s \in (0.5, 1)$ , where  $C > 0$  is a constant.

### 2.1. The strategy for proportion estimation

Let  $\mathbf{z} = (z_1, \dots, z_m)^\top$  and  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)^\top$ . Denote by  $F_{\mu_i}(\cdot)$  the CDF of  $z_i$  for  $i \in \{1, \dots, m\}$  and suppose each  $F_{\mu_i}$  with  $i \in \{0, 1, \dots, m\}$  is a member of a set  $\mathcal{F}$  of CDFs such that  $\mathcal{F} = \{F_\mu : \mu \in U\}$  for some non-empty  $U$  in  $\mathbb{R}$ . For the rest of the paper, we assume that each  $F_\mu$  is uniquely determined by  $\mu$  and that  $U$  has a non-empty interior.

To illustrate the intuitions behind Jin’s strategy, we consider the setting where each  $z_i$  is Normally distributed with mean  $\mu_i$  and standard deviation 1 (denoted by  $z_i \sim \mathcal{N}(\mu_i, 1)$ ) and  $\mu_0 = 0$ . So,  $\pi_{1,m} = m^{-1} \sum_{i=1}^m (1 - 1_{\{\mu_i=0\}})$ . If we can construct a function  $\hat{\psi}(t, \mu)$  such that  $\hat{\psi}(t, 0) = 1$  for all  $t \in \mathbb{R}$  and  $\lim_{t \rightarrow \infty} \hat{\psi}(t, \mu) = 0$  for any  $\mu \neq 0$ , then, for any fixed  $\boldsymbol{\mu}$  and  $m$ ,  $1_{\{\mu_i=0\}} = \hat{\psi}(t, 0)$  when  $\mu_i = 0$  and  $1_{\{\mu_i \neq 0\}} = 1 - \lim_{t \rightarrow \infty} \hat{\psi}(t, \mu_i)$  when  $\mu_i \neq 0$ . In other words,

$$\pi_{1,m} = \lim_{t \rightarrow \infty} m^{-1} \sum_{i=1}^m \left\{ 1 - \hat{\psi}(t, \mu_i) \right\} \text{ for any fixed } \boldsymbol{\mu} \text{ and } m.$$

One way to construct  $\hat{\psi}$  is to utilize the Riemann–Lebesgue lemma (see, e.g., [12]) and represent  $\hat{\psi}$  as an integral. For example, given any probability density function  $\tilde{\omega}$  on  $[-1, 1]$ , setting

$$\hat{\psi}(t, \mu) = \int \tilde{\omega}(s) \cos(t\mu s) ds$$

immediately gives the desired  $\hat{\psi}$ . If  $\tilde{\omega}$  is also even, then  $\hat{\psi}$  is the Fourier transform of  $\tilde{\omega}$  evaluated at  $t\mu$ . Notice that  $\hat{\psi}$  is a deterministic function and that  $\tilde{\omega}$  is referred to as an “averaging function” in Section 3. Even though a  $\hat{\psi}$  has been found, it cannot be used to estimate  $\pi_{1,m}$  since  $\hat{\psi}$  is not a function of any  $z_i$ . So, the next step is to connect  $\hat{\psi}(t, \mu_i)$  with  $z_i$  probabilistically via a function  $\hat{K}(x, t)$ , so that  $\mathbb{E} \left\{ \hat{K}(z_i, t) \right\} = \hat{\psi}(t, \mu_i)$  for each  $t$  and  $i$ . Once such a  $\hat{K}$  is found, then  $\pi_{1,m}^\circ(t) = m^{-1} \sum_{i=1}^m \left\{ 1 - \hat{K}(z_i, t) \right\}$  serves as an estimate of  $\pi_{1,m}$  and can be very accurate when  $t$  is large. Specifically, setting

$$\hat{K}(x, t) = \int \tilde{\omega}(s) \exp(2^{-1}t^2s^2) \cos(txs) ds$$

gives  $\mathbb{E} \left\{ \hat{K}(z_i, t) \right\} = \hat{\psi}(t, \mu_i)$  whenever  $z_i \sim \mathcal{N}(\mu_i, 1)$ , and  $\hat{\psi}(t, \mu) = \left\{ \hat{K}(\cdot, t) * \phi \right\}(\mu)$ , where  $\phi$  is the standard Normal density and  $*$  denotes convolution. This fact can be found in [20] or [7]. Finally, to show the consistency of  $\pi_{1,m}^\circ(t)$ , we only need to control the difference  $\pi_{1,m}^\circ(t) - \mathbb{E} \left\{ \pi_{1,m}^\circ(t) \right\}$  as both  $t$  and  $m$  tend to  $\infty$ .

Now we state the Strategy below. For each fixed  $\mu \in U$ , if we can approximate the indicator function  $1_{\{\mu \neq \mu_0\}}$  by a function  $\psi(t, \mu; \mu_0)$  with  $t \in \mathbb{R}$  satisfying  $\lim_{t \rightarrow \infty} \psi(t, \mu_0; \mu_0) = 1$  and  $\lim_{t \rightarrow \infty} \psi(t, \mu; \mu_0) = 0$  for  $\mu \neq \mu_0$ , then the “phase function”

$$\varphi_m(t, \boldsymbol{\mu}) = \frac{1}{m} \sum_{i=1}^m \{1 - \psi(t, \mu_i; \mu_0)\}$$

satisfies  $\lim_{t \rightarrow \infty} \varphi_m(t, \boldsymbol{\mu}) = \pi_{1,m}$  for any fixed  $m$  and  $\boldsymbol{\mu}$  and provides the “Oracle”  $\Lambda_m(\boldsymbol{\mu}) = \lim_{t \rightarrow \infty} \varphi_m(t, \boldsymbol{\mu})$ . Further, if we can find a function  $K : \mathbb{R}^2 \rightarrow \mathbb{R}$  that does not depend on any  $\mu \neq \mu_0$  and satisfies the Lebesgue–Stieltjes integral equation

$$\psi(t, \mu; \mu_0) = \int K(t, x; \mu_0) dF_\mu(x), \quad (1)$$

then the “empirical phase function”

$$\hat{\varphi}_m(t, \mathbf{z}) = \frac{1}{m} \sum_{i=1}^m \{1 - K(t, z_i; \mu_0)\}$$

satisfies  $\mathbb{E} \left\{ \hat{\varphi}_m(t, \mathbf{z}) \right\} = \varphi_m(t, \boldsymbol{\mu})$  for any fixed  $m, t$  and  $\boldsymbol{\mu}$ . Namely,  $\hat{\varphi}_m(t, \mathbf{z})$  is an unbiased estimator of  $\varphi_m(t, \boldsymbol{\mu})$ . By the laws of large numbers,  $\hat{\varphi}_m(t, \mathbf{z})$  can be close to  $\varphi_m(t, \boldsymbol{\mu})$  for a fixed  $t$  when  $m$  is large. When the difference

$$e_m(t) = \left| \hat{\varphi}_m(t, \mathbf{z}) - \varphi_m(t, \boldsymbol{\mu}) \right| \quad (2)$$

is suitably small for large  $t$ ,  $\hat{\varphi}_m(t, \mathbf{z})$  will accurately estimate  $\pi_{1,m}$ . Since  $\varphi_m(t, \boldsymbol{\mu}) = \pi_{1,m}$  or  $\hat{\varphi}_m(t, \mathbf{z}) = \pi_{1,m}$  rarely happens,  $\hat{\varphi}_m(t, \mathbf{z})$  usually employs a monotone increasing sequence  $\{t_m\}_{m \geq 1}$  such that  $\lim_{m \rightarrow \infty} t_m = \infty$  in order to achieve consistency in the sense that

$$\Pr \left\{ \left| \frac{\hat{\varphi}_m(t_m, \mathbf{z})}{\pi_{1,m}} - 1 \right| \rightarrow 0 \right\} \rightarrow 1 \quad \text{as } m \rightarrow \infty. \quad (3)$$

So, the intrinsic speed for  $\hat{\varphi}_m(t_m, \mathbf{z})$  to achieve consistency is better represented by  $t_m$ , and we will use  $t_m$  as the “speed of convergence” of  $\hat{\varphi}_m(t_m, \mathbf{z})$ . Throughout the paper, consistency of a proportion estimator is defined via (3) to accommodate the scenario  $\lim_{m \rightarrow \infty} \pi_{1,m} = 0$ .

By duality,  $\varphi_m^*(t, \boldsymbol{\mu}) = 1 - \varphi_m(t, \boldsymbol{\mu})$  is the oracle for which  $\pi_{0,m} = \lim_{t \rightarrow \infty} \varphi_m^*(t, \boldsymbol{\mu})$  for any fixed  $m$  and  $\boldsymbol{\mu}$ ,  $\hat{\varphi}_m^* = 1 - \hat{\varphi}_m(t, \mathbf{z})$  satisfies  $\mathbb{E} \{ \hat{\varphi}_m^*(t, \mathbf{z}) \} = \varphi_m^*(t, \boldsymbol{\mu})$  for any fixed  $m, t$  and  $\boldsymbol{\mu}$ , and  $\hat{\varphi}_m^*(t, \mathbf{z})$  will accurately estimate  $\pi_{0,m}$  when  $e_m(t)$  is suitably small for large  $t$ . Further, the stochastic oscillations of  $\hat{\varphi}_m^*(t, \mathbf{z})$  and  $\hat{\varphi}_m(t, \mathbf{z})$  are the same and is quantified by  $e_m(t)$ .

We remark on the differences between the Strategy and Jin’s strategy. The latter in our notations sets  $\mu_0 = 0$ , requires  $\psi(t, 0; 0) = 1$  for all  $t$ , requires location-shift families and uses Fourier transform to construct  $K$  and  $\psi$ , deals with distributions whose means are equal to their medians, and intends to have

$$1 \geq \psi(t, \boldsymbol{\mu}; 0) \geq 0 \quad \text{for all } \boldsymbol{\mu} \text{ and } t. \quad (4)$$

It is not hard to see from the proof of Lemma 7.1 of [20] that (4) may not be achievable for non-location-shift families. Further, Jin’s construction of  $\psi$  is a special case of solving (1). Moreover, it is easier to solve (1) for  $K : \mathbb{R}^2 \rightarrow \mathbb{C}$  in the complex domain. However, a real-valued  $K$  is preferred for applications in statistics.

### 3. Construction I

In this section, we present the construction, referred to as “Construction I”, when the set of characteristic functions (CFs) of the CDFs of  $\{z_i\}_{i=1}^m$  are of Riemann–Lebesgue type (see Definition 1). This construction essentially depends on generalizations of the Riemann–Lebesgue lemma (“RL lemma”) and subsumes that by [20] for Gaussian family.

Recall the family of CDFs  $\mathcal{F} = \{F_\mu : \mu \in U\}$  and let  $\hat{F}_\mu(t) = \int e^{itx} dF_\mu(x)$  be the CF of  $F_\mu$  where  $\iota = \sqrt{-1}$ . Let  $r_\mu$  be the modulus of  $\hat{F}_\mu$ . Then  $\hat{F}_\mu = r_\mu e^{ih_\mu}$ , where  $h_\mu$  is the principal value of the argument that ranges in  $(-\pi, \pi]$ . Further,  $h_\mu$  restricted to the non-empty open interval  $(-\tau_\mu, \tau_\mu)$  with  $h_\mu(0) = 0$  is uniquely defined, continuous and odd, where  $\tau_\mu = \inf \{t > 0 : \hat{F}_\mu(t) = 0\} > 0$ ; see [30] for a positive lower bound for  $\tau_\mu$ . Note that  $\hat{F}_\mu$  has no real zeros if and only if  $\tau_\mu = \infty$ .

**Definition 1.** If

$$\{t \in \mathbb{R} : \hat{F}_{\mu_0}(t) = 0\} = \emptyset \quad (5)$$

and, for each  $\mu \in U \setminus \{\mu_0\}$

$$\sup_{t \in \mathbb{R}} \frac{r_\mu(t)}{r_{\mu_0}(t)} < \infty \quad (6)$$

and

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_{[-t, t]} \frac{\hat{F}_\mu(y)}{\hat{F}_{\mu_0}(y)} dy = 0, \quad (7)$$

then  $\hat{\mathcal{F}} = \{\hat{F}_\mu : \mu \in U\}$  is said to be of “Riemann–Lebesgue type (RL type)” (at  $\mu_0$  on  $U$ ).

In Definition 1, condition (5) requires that  $\hat{F}_{\mu_0}$  have no real zeros, (6) that  $r_\mu$  with  $\mu \in U \setminus \{\mu_0\}$  approximately be of the same order as  $r_{\mu_0}$ , and (7) forces the “mean value” of  $\hat{F}_\mu \hat{F}_{\mu_0}^{-1}$  to converge to zero. Condition (5) excludes the Pólya-type CF  $\hat{F}(t) = \{1 - |t|\} 1_{\{|t| \leq 1\}}$  but holds for an infinitely divisible CF, (6) usually cannot be relaxed to be  $r_\mu r_{\mu_0}^{-1} \in L^1(\mathbb{R})$  as seen from Lemma 1 for location-shift families, and (7) induces a generalization of the RL lemma given by [12] (which enables a construction via Fourier transform).

With a Fourier transform based construction comes the subtle issue of determining an “averaging function”  $\omega$  that helps invoke the RL lemma and facilitates easy numerical implementation of the resulting proportion estimator. Indeed, a carefully chosen  $\omega$  will greatly simplify the construction and induce agreeable proportion estimators. We adapt from [20] the concept of a “good”  $\omega$ :

**Definition 2.** If a function  $\omega : [-1, 1] \rightarrow \mathbb{R}$  is non-negative and bounded such that  $\int_{[-1, 1]} \omega(s) ds = 1$ , then it is called “admissible”. If additionally  $\omega$  is even on  $[-1, 1]$  and continuous on  $(-1, 1)$ , then it is called “eligible”. If  $\omega$  is eligible and  $\omega(t) \leq \tilde{\omega}(1 - t)$  for all  $t \in (0, 1)$  for some convex, super-additive function  $\tilde{\omega}$  over  $(0, 1)$ , then it is called “good”.



The definition above includes the end points  $\{-1, 1\}$  of the compact interval  $[-1, 1]$  to tame  $\omega$  at these points. For example, the triangular density  $\omega(s) = \max\{1 - |s|, 0\}$  is good and its CF is

$$\hat{\omega}(t) = \frac{2\{1 - \cos(t)\}}{t^2} 1_{\{t \neq 0\}} + 1_{\{t=0\}}$$

as discussed by [20]. Since  $r_\mu(t)$  is even and uniformly continuous in  $t$ , an even  $\omega$  matches  $r_\mu$  and can induce an even  $K$  as a function of  $t$ . Unless otherwise noted,  $\omega$  in this work is always admissible.

**Theorem 1.** Let  $\hat{\mathcal{F}}$  be of RL type and define  $K : \mathbb{R}^2 \rightarrow \mathbb{R}$  as

$$K(t, x; \mu_0) = \int_{[-1, 1]} \frac{\omega(s) \cos\{tsx - h_{\mu_0}(ts)\}}{r_{\mu_0}(ts)} ds. \quad (8)$$

Then  $\psi(t, \mu; \mu_0)$  in (1) satisfies the following:

1.  $\psi : \mathbb{R} \times U \rightarrow \mathbb{R}$  with

$$\psi(t, \mu; \mu_0) = \int_{[-1, 1]} \omega(s) \frac{r_\mu(ts)}{r_{\mu_0}(ts)} \cos\{h_\mu(ts) - h_{\mu_0}(ts)\} ds. \quad (9)$$

2.  $\psi(t, \mu_0; \mu_0) = 1$  for all  $t$ , and  $\lim_{t \rightarrow \infty} \psi(t, \mu; \mu_0) = 0$  for each  $\mu \in U$  such that  $\mu \neq \mu_0$ .

For random variables with RL type CFs in general, it is hard to ensure (4), i.e.,  $1 \geq \psi(t, \mu; 0) \geq 0$  for all  $\mu$  and  $t$ , since we do not have sufficient information on the phase  $h_\mu$ . However, for certain location-shift families, (4) holds when  $\omega$  is good; see Corollary 2. Under slightly stronger conditions, we have:

**Corollary 1.** Assume that (5) and (6) hold. If  $r_\mu/r_{\mu_0} \in L^1(\mathbb{R})$ , then the conclusions of Theorem 1 hold.

When each  $F_\mu$  has a density with respect to  $\nu$ , the condition  $r_\mu r_{\mu_0}^{-1} \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$  in Corollary 1 can be fairly strong since it forces  $\lim_{t \rightarrow \infty} r_\mu(t)/r_{\mu_0}(t) = 0$ . Unfortunately, Corollary 1 is not applicable to location-shift families since for these families  $r_\mu(t)/r_{\mu_0}(t) \equiv 1$  for all  $\mu \in U$  whenever  $r_{\mu_0}(t) \neq 0$ ; see Lemma 1. In contrast, Theorem 1 is as we explain next.

Recall the definition of the location-shift family, i.e.,  $\mathcal{F} = \{F_\mu : \mu \in U\}$  is a location-shift family if and only if  $z + \mu'$  has CDF  $F_{\mu+\mu'}$  whenever  $z$  has CDF  $F_\mu$  for  $\mu, \mu + \mu' \in U$ .

**Lemma 1.** Suppose  $\mathcal{F}$  is a location-shift family. Then,  $r_\mu$  does not depend on  $\mu$  and  $h_\mu(t) = t\mu'$ , where  $\mu' = \mu - \mu_0$ . If in addition  $\hat{F}_{\mu_0}(t) \neq 0$  for all  $t \in \mathbb{R}$ , then  $\hat{\mathcal{F}}$  is of RL type.

Lemma 1 implies that very likely the set of location-shift family distributions is a subset of the set of distributions with RL type CFs. However, verifying if  $\hat{F}_{\mu_0}$  for a location-shift family has any real zeros is quite difficult in general. Nonetheless, with Lemma 1, Theorem 1 implies:

**Corollary 2.** If  $\mathcal{F}$  is a location-shift family for which (5) holds, then

$$\psi(t, \mu; \mu_0) = \int K(t, y + \mu - \mu_0; \mu_0) dF_{\mu_0}(y) = \int_{[-1, 1]} \omega(s) \cos\{ts(\mu - \mu_0)\} ds. \quad (10)$$

If additionally  $\omega$  is good, then (4) holds, i.e.,  $1 \geq \psi(t, \mu; \mu_0) \geq 0$  for all  $\mu$  and  $t$ .

The identity (10) in Corollary 2 asserts that, for a location-shift family,  $\psi$  in (1) reduces to a “convolution”, and it is the intrinsic mechanism behind the Fourier transform based construction. When  $\mu_0 = 0$ , from Corollary 2 we directly have

$$\psi(t, \mu; 0) = \int K(t, x + \mu; 0) dF_0(x) = \int_{[-1, 1]} \omega(s) \cos(ts\mu) ds,$$

recovering the construction by [20] for Gaussian and Laplace families. Examples for which (10) holds are given in Section 3.1.

### 3.1. Construction I: some examples

We provide some examples from Construction I. They are all formed out of location-shift families by allowing the location parameter  $\mu$  to vary but fixing the scale parameter  $\sigma$ , and are infinitely divisible so that none of their CFs has any real zero; see [15,29] and [34] for details on this. Note however that the Poisson family as an example given in Section 4.3 is infinitely divisible but does not have an RL type CF. There are certainly other examples that Construction I applies to. But we will not attempt to exhaust them.

**Example 1.** Gaussian family  $\mathcal{N}(\mu, \sigma^2)$  with mean  $\mu$  and standard deviation  $\sigma > 0$ , for which

$$\frac{dF_\mu}{dv}(x) = f_\mu(x) = \left(\sqrt{2\pi}\sigma\right)^{-1} \exp\left\{-2^{-1}\sigma^{-2}(x-\mu)^2\right\}.$$

The CF of  $f_\mu$  is  $\hat{f}_\mu(t) = \exp(it\mu) \exp(-2^{-1}t^2\sigma^2)$ . Here  $r_\mu^{-1}(t) = \exp(2^{-1}t^2\sigma^2)$ . Therefore,

$$K(t, x; \mu_0) = \int_{[-1,1]} r_\mu^{-1}(ts) \omega(s) \exp\{its(x - \mu_0)\} ds \quad (11)$$

and (10) holds. When  $\mu_0 = 0$  and  $\sigma = 1$ , this construction reduces to that in Section 2.1 of [20].

**Example 2.** Laplace family  $\text{Laplace}(\mu, 2\sigma^2)$  with mean  $\mu$  and standard deviation  $\sqrt{2}\sigma > 0$  for which

$$\frac{dF_\mu}{dv}(x) = f_\mu(x) = \frac{1}{2\sigma} \exp(-\sigma^{-1}|x - \mu|)$$

and the CF is  $\hat{f}_\mu(t) = (1 + \sigma^2 t^2)^{-1} \exp(it\mu)$ . Therefore, (11) and (10) hold with  $r_\mu^{-1}(t) = 1 + \sigma^2 t^2$ . When  $\mu_0 = 0$  and  $\sigma = 1$ , this construction reduces to that in Section 7.1 of [20].

**Example 3.** Logistic family  $\text{Logistic}(\mu, \sigma)$  with mean  $\mu$  and scale parameter  $\sigma > 0$ , for which

$$\frac{dF_\mu}{dv}(x) = f_\mu(x) = \frac{1}{4\sigma} \operatorname{sech}^2\left(\frac{x - \mu}{2\sigma}\right)$$

and the CF of  $f_\mu$  is  $\hat{f}_\mu(t) = \pi\sigma t \{\sinh(\pi\sigma t)\}^{-1} \exp(it\mu)$ . Therefore, (11) and (10) hold with  $r_\mu^{-1}(t) = (\pi\sigma t)^{-1} \sinh(\pi\sigma t) \sim (2\pi\sigma t)^{-1} e^{\pi\sigma t}$  as  $t \rightarrow \infty$ .

**Example 4.** Cauchy family  $\text{Cauchy}(\mu, \sigma)$  with median  $\mu$  and scale parameter  $\sigma > 0$ , for which

$$\frac{dF_\mu}{dv}(x) = f_\mu(x) = \frac{1}{\pi\sigma} \frac{\sigma^2}{(x - \mu)^2 + \sigma^2}$$

and the CF is  $\hat{f}_\mu(t) = \exp(-\sigma|t|) \exp(it\mu)$ . Therefore, (11) and (10) hold with  $r_\mu^{-1}(t) = \exp(\sigma|t|)$ . Note that the Cauchy family does not have a finite first-order absolute moment.

**Example 5.** Hyperbolic Secant family  $\text{HSecant}(\mu, \sigma)$  with mean  $\mu$  and scale parameter  $\sigma > 0$ , for which

$$\frac{dF_\mu}{dv}(x) = f_\mu(x) = \frac{1}{2\sigma} \frac{1}{\cosh\left(\pi \frac{x - \mu}{\sigma}\right)};$$

see, e.g., Chapter 1 of [15]. The identity

$$\int_{-\infty}^{+\infty} e^{tx} \frac{dx}{\pi \cosh(x)} = \cosh(2^{-1}\pi t),$$

implies  $\hat{f}_\mu(t) = \sigma^{-1} \exp(-it\mu\sigma^{-1}) \operatorname{sech}(t\sigma^{-1})$ . Therefore, (11) and (10) hold with  $r_\mu^{-1}(t) = \sigma \cosh(t\sigma^{-1}) \sim 2^{-1}\sigma \exp(\sigma^{-1}t)$  as  $t \rightarrow \infty$ . The Hyperbolic Secant family has been used to model stock indices and exchange rates [15] or status of coronary heart disease [43].

### 3.2. Construction I: uniform consistency and speed of convergence

The performance of the estimator  $\hat{\varphi}_m$  depends on how accurately it approximates  $\varphi_m$ , the oracle that knows the true value  $\pi_{1,m}$  as  $t \rightarrow \infty$ . Specifically, the smaller  $e_m(t)$  (defined by (2)) is when  $t$  is large, the more accurately  $\hat{\varphi}_m(t; \mathbf{z})$  estimates  $\pi_{1,m}$ . Two key factors that affect  $e_m$  are: (i) the magnitude of the reciprocal of the modulus,  $r_{\mu_0}^{-1}(t)$ , which appears as a scaling factor in the integrand in the definition of  $K$  in (8), and (ii) the magnitudes of the  $\mu_i$ 's and the variabilities of  $z_i$ 's. As will be shown by Theorem 2, for independent  $\{z_i\}_{i=1}^m$ , the oscillation of  $e_m(t)$  depends mainly on  $r_{\mu_0}^{-1}(t)$  due to the concentration of measure for independent, uniformly bounded random variables and their transforms by Lipschitz functions, whereas the consistency of  $\hat{\varphi}_m(t; \mathbf{z})$  depends also on the magnitudes of  $\pi_{1,m}$  and  $\mu_i$ 's that affect how accurate the oracle is when  $t$  is large.

**Theorem 2.** Let  $a(t; \mu_0) = \int_{[-1,1]} r_{\mu_0}^{-1}(ts) ds$  for  $t \in \mathbb{R}$ . If  $\hat{\mathcal{F}}$  is of RL type and  $\{z_i\}_{i=1}^m$  are independent, then

$$\mathbb{V}\left\{\left|\hat{\varphi}_m(t, \mathbf{z}) - \varphi_m(t, \boldsymbol{\mu})\right|\right\} \leq \frac{1}{m} \|\omega\|_\infty^2 a^2(t; \mu_0). \quad (12)$$



Further, for any fixed  $\lambda > 0$ , with probability at least  $1 - 2 \exp(-2^{-1}\lambda^2)$ ,

$$|\hat{\varphi}_m(t, \mathbf{z}) - \varphi_m(t, \boldsymbol{\mu})| \leq \frac{\lambda \|\omega\|_\infty a(t; \mu_0)}{\sqrt{m}}. \quad (13)$$

If there are positive sequences  $\{u_m\}_{m \geq 1}$ ,  $\{\lambda_m\}_{m \geq 1}$  and  $\{t_m\}_{m \geq 1}$  such that

$$\lim_{m \rightarrow \infty} \sup \{\psi(t, \mu; \mu_0) : (t, |\mu|) \in [t_m, \infty) \times [u_m, \infty)\} = 0 \quad (14)$$

and

$$\lim_{m \rightarrow \infty} \frac{\lambda_m a(t_m; \mu_0)}{\pi_{1,m} \sqrt{m}} = 0 \text{ and } \lim_{m \rightarrow \infty} \exp(-2^{-1}\lambda_m^2) = 0, \quad (15)$$

then

$$\Pr\{|\pi_{1,m}^{-1} \hat{\varphi}_m(t_m, \mathbf{z}) - 1| \rightarrow 0\} \rightarrow 1 \quad (16)$$

whenever  $\{\mu_i : i \in I_{1,m}\} \subseteq [u_m, \infty)$ .

**Theorem 2** bounds the variance of  $\hat{\varphi}_m(t, \mathbf{z})$ , captures the key ingredients needed for and the essence of proving the consistency of a proportion estimator based on Construction I, and shows that such a proportion estimator is consistent as long as  $\lambda_m \rightarrow \infty$ ,  $\lambda_m \pi_{1,m}^{-1} a(t_m; \mu_0)$  is of smaller order than  $\sqrt{m}$ , and each  $\psi(t_m, \mu_i; \mu_0)$  with  $i \in I_{1,m}$  is negligible for a sequence  $t_m \rightarrow \infty$ . Using **Theorem 2**, we can characterize the consistency of the estimator  $\hat{\varphi}_m(t, \mathbf{z})$  for each example given in Section 3.1 as follows:

**Corollary 3.** Consider  $\hat{\varphi}(t_m, \mathbf{z})$  from Construction I. Let  $\{z_i\}_{i=1}^m$  be independent and  $\{t_m : m \geq 1\}$  a positive sequence such that  $\lim_{m \rightarrow \infty} t_m = \infty$ . Assume  $\min_{i \in I_{1,m}} |\mu_i - \mu_0| \geq (t_m)^{-1} \ln \ln m$ . Then  $\Pr\{|\pi_{1,m}^{-1} \hat{\varphi}_m(t_m, \mathbf{z}) - 1| \rightarrow 0\} \rightarrow 1$  holds

- for Gaussian, Hyperbolic Secant, Logistic and Cauchy family respectively when  $t_m = \sigma^{-1} \sqrt{2\gamma \ln m}$ ,  $t_m = \sigma \gamma \ln m$ ,  $t_m = (\sigma\pi)^{-1} \gamma \ln m$  and  $t_m = \sigma^{-1} \gamma \ln m$ ,  $\lambda_m = o(t_m)$ ,  $\lambda_m \rightarrow \infty$  and  $\pi_{1,m} \geq C m^{\gamma-0.5}$  with  $\gamma \in (0, 0.5]$ ;
- for Laplace family when  $t_m = \ln m$ ,  $\lambda_m = O(t_m)$ ,  $\lambda_m \rightarrow \infty$  and  $\pi_{1,m} \geq C m^{-\gamma}$  with  $\gamma \in [0, 0.5)$ ,

where  $C > 0$  can be any constant for which  $\pi_{1,m} \in (0, 1]$  as  $\gamma$  varies in its designated range respectively.

To characterize if an estimator  $\hat{\varphi}_m(t, \mathbf{z})$  is uniformly consistent with respect to  $\pi_{1,m}$  and  $t$ , we introduce the following definition:

**Definition 3.** Given a family  $\mathcal{F}$ , the sequence of sets  $\mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F}) \subseteq \mathbb{R}^m \times \mathbb{R}$  for each  $m \in \mathbb{N}_+$  is called a “uniform consistency class” for the estimator  $\hat{\varphi}_m(t, \mathbf{z})$  if

$$\Pr\left\{\sup_{\boldsymbol{\mu} \in \mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F})} \left|\pi_{1,m}^{-1} \sup_{t \in \mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F})} \hat{\varphi}_m(t, \mathbf{z}) - 1\right| \rightarrow 0\right\} \rightarrow 1. \quad (17)$$

If (17) holds and the  $t$ -section of  $\mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F})$  (that is a subset of  $\mathbb{R}^m$  containing  $\boldsymbol{\mu}$ ) does not converge to the empty set in  $\mathbb{R}^m$  as  $m \rightarrow \infty$ , then  $\hat{\varphi}_m(t, \mathbf{z})$  is said to be “uniformly consistent”. If further the  $\boldsymbol{\mu}$ -section of  $\mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F})$  (that is a subset of  $\mathbb{R}$  containing  $t$ ) contains a connected subset  $G_m \subseteq \mathbb{R}$  such that  $\lim_{m \rightarrow \infty} \nu(G_m) = \infty$ , then  $\hat{\varphi}_m(t, \mathbf{z})$  is said to be “uniformly consistent in frequency domain”.

Now we discuss uniform consistency in frequency domain of an estimator from Construction I. Define

$$\mathcal{B}_m(\rho) = \left\{\boldsymbol{\mu} \in \mathbb{R}^m : m^{-1} \sum_{i=1}^m |\mu_i - \mu_0| \leq \rho\right\} \text{ for some } \rho > 0$$

and  $u_m = \min\{|\mu_j - \mu_0| : \mu_j \neq \mu_0\}$ .

**Theorem 3.** Assume that  $\mathcal{F}$  is a location-shift family for which (5) holds and  $\int |x|^2 dF_\mu(x) < \infty$  for each  $\mu \in U$ . If  $\sup_{y \in \mathbb{R}} \left|\frac{d}{dy} h_{\mu_0}(y)\right| = C_{\mu_0} < \infty$ , then for the estimator  $\hat{\varphi}_m(t, \mathbf{z})$  from Construction I, a uniform consistency class is

$$\mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F}) = \left\{\begin{array}{l} q\gamma' > \vartheta > 2^{-1}, \gamma' > 0, \gamma'' > 0, 0 \leq \vartheta' < \vartheta - 1/2, \\ R_m(\rho) = O(m^{\vartheta'}), \tau_m \leq \gamma_m, u_m \geq \frac{\ln \ln m}{\gamma'' \tau_m}, \\ t \in [0, \tau_m], \lim_{m \rightarrow \infty} \pi_{1,m}^{-1} \gamma(q, \tau_m, \gamma_m, r_{\mu_0}) = 0 \end{array}\right\}$$

where  $q, \gamma', \gamma'', \vartheta$  and  $\vartheta'$  are constants,  $R_m(\rho) = 2 \int |x| dF_{\mu_0}(x) + 2\rho + 2C_{\mu_0}$ ,  $\gamma_m = \gamma' \ln m$  and

$$\gamma(q, \tau_m, \gamma_m, r_{\mu_0}) = \frac{2 \|\omega\|_\infty \sqrt{2q\gamma_m}}{\sqrt{m}} \sup_{t \in [0, \tau_m]} \int_{[0,1]} \frac{ds}{r_{\mu_0}(ts)}.$$

Moreover, for all sufficiently large  $m$ ,

$$\sup_{\mu \in \mathcal{B}_m(\rho)} \sup_{t \in [0, \tau_m]} |\hat{\varphi}_m(t, \mathbf{z}) - \varphi_m(t, \boldsymbol{\mu})| \leq \Upsilon(q, \tau_m, \gamma_m, r_{\mu_0}) \quad (18)$$

holds with probability at least  $1 - o(1)$ .

Several remarks on Theorem 3 are ready to be stated. Firstly, since Theorem 3 requires the random variables to have finite absolute second-order moments, it may not apply to location-shift families that do not have finite first-order absolute moments. Secondly, even though Theorem 3 potentially allows for many possible choices of  $t$  in  $\hat{\varphi}_m(t, \mathbf{z})$ , we should choose  $\tau_m$  such that  $\tau_m \rightarrow \infty$  as fast as possible so that  $\pi_{1,m}^{-1} \hat{\varphi}_m(\tau_m, \mathbf{z}) \rightarrow 1$  as fast as possible. Thirdly, compared to Theorems 1.4 and 1.5 of [20] where  $\mathcal{B}_m(\rho)$  is for a fixed  $\rho$  for Gaussian family, we allow  $\rho \rightarrow \infty$  for location-shift families. Fourthly, the bound (18) together with 3 imply that, when other things are kept fixed, the larger  $r_{\mu}^{-1}$  is, the slower  $\pi_{1,m}^{-1} \hat{\varphi}_m(t_m, \mathbf{z}) \rightarrow 1$ . This has been observed by [20] for the Gaussian and Laplace families since  $r_{\mu}^{-1}(t)$  for the former is much larger than the latter when  $t$  is large. Fifthly, for location-shift families the constants  $q, \gamma', \gamma''$  and  $\vartheta$  can be specified by a user. If additionally  $\vartheta' = 0$ , then  $\mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F})$  is fully data-adaptive and depends only on  $u_m = \min_{i \in I_{1,m}} |\mu_i - \mu_0|$ ; see also Corollary 4 on this for examples given in Section 3.1.

Let  $\mathcal{F}$  and  $\tilde{\mathcal{F}}$  be two location-shift families that have RL type CFs and are determined by the same set of parameters, and  $r_{\mu_0}$  and  $\tilde{r}_{\mu_0}$  respectively the moduli of the CFs of  $F_{\mu_0} \in \mathcal{F}$  and  $\tilde{F}_{\mu_0} \in \tilde{\mathcal{F}}$ . For the estimator  $\hat{\varphi}_m(t, \mathbf{z})$  under the same settings and under the conditions stated in Theorem 3, consider the two uniform consistency classes  $\mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F})$  and  $\mathcal{Q}_m(\boldsymbol{\mu}, t; \tilde{\mathcal{F}})$  for  $\hat{\varphi}_m(t, \mathbf{z})$  that have the same constants  $q, \gamma', \vartheta$  and  $\gamma''$ . Then Theorem 3 implies that

$$r_{\mu_0}(t) \leq \tilde{r}_{\mu_0}(t) \text{ for all large } t > 0 \implies \mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F}) \supseteq \mathcal{Q}_m(\boldsymbol{\mu}, t; \tilde{\mathcal{F}}). \quad (19)$$

In other words, for location-shift families that have RL type CFs, the uniform consistency classes for estimators from Construction I can be reversely ordered via set inclusion by the magnitudes of the associated moduli when other conditions are the same.

Another consequence of Theorem 3 is as follows. If we fix  $q, \gamma', \gamma'', \vartheta$  and  $\vartheta'$ , then  $r_{\mu_0}$  determines  $\mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F})$ . In particular, if  $F_{\mu}$  has a density with respect to  $\nu$ , then  $\lim_{t \rightarrow \infty} r_{\mu}(t) = 0$  must hold, which forces 3 to give

$$\Upsilon(q, \tau_m, \gamma' \ln m, r_{\mu_0}) \geq \frac{C\sqrt{q\gamma' \ln m}}{\sqrt{m}} \text{ as } m \rightarrow \infty. \quad (20)$$

Since Hoeffding's inequality, used in the proof of Theorem 3, is asymptotically optimal for independent, uniformly almost surely bounded random variables (see [11]) that are induced by Construction I, Theorem 3 and (20) together imply that, when Construction I is applied to location-shift families with absolutely continuous CDFs, a uniform consistency class for the corresponding proportion estimator is unlikely able to contain any  $\pi_{1,m} \in (0, m^{-0.5}]$ . Comparing this with the conclusions of Theorem 2 and Corollary 3, we see that a sacrifice to achieve uniform consistency in frequency domain is the reduction of the range for  $\pi_{1,m}$  for which  $\hat{\varphi}_m(t, \mathbf{z})$  can be consistent.

Finally, Theorem 3 gives

**Corollary 4.** When  $\{z_i\}_{i=1}^m$  are independent, the following results hold for  $\hat{\varphi}_m(t, \mathbf{z})$ :

(i) For Gaussian family:  $q\sigma^{-1} > \vartheta > 2^{-1}$ ,  $0 \leq \vartheta' < \vartheta - 1/2$ ,  $\gamma_m = \sigma^{-1} \ln m$  and

$$\mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F}) = \left\{ \begin{array}{l} R_m(\rho) = O(m^{\vartheta'}), u_m \geq \frac{\ln \ln m}{\sqrt{2 \ln m}}, 0 < \gamma \leq 0.5, \\ \pi_{1,m} \geq Cm^{\gamma-0.5}, t \in [0, \sigma^{-1} \sqrt{2\gamma \ln m}] \end{array} \right\}; \quad (21)$$

the fastest speed of convergence is  $\sqrt{\ln m}$ , achieved when  $\liminf_{m \rightarrow \infty} \pi_{1,m} > 0$ ;

(ii) For Laplace family:  $q > \vartheta > 2^{-1}$ ,  $0 \leq \vartheta' < \vartheta - 1/2$ ,  $\gamma_m = \ln m$  and

$$\mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F}) = \left\{ \begin{array}{l} R_m(\rho) = O(m^{\vartheta'}), u_m \geq \frac{\ln \ln m}{\ln m}, 0 \leq \gamma < 1/2, \\ \pi_{1,m} \geq Cm^{-\gamma}, t \in [0, \ln m] \end{array} \right\};$$

(iii) For Hyperbolic Secant family:  $q\sigma > \vartheta > 2^{-1}$ ,  $\gamma_m = \sigma \ln m$ ,  $0 \leq \vartheta' < \vartheta - 1/2$  and

$$\mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F}) = \left\{ \begin{array}{l} R_m(\rho) = O(m^{\vartheta'}), u_m \geq \frac{\ln \ln m}{\ln m}, 0 < \gamma < 1/2, \\ \pi_{1,m} \geq Cm^{\gamma-0.5}, t \in [0, \sigma \gamma \ln m] \end{array} \right\};$$

(iv) For Logistic family:  $q(\sigma\pi)^{-1} > \vartheta > 2^{-1}$ ,  $\gamma_m = (\sigma\pi)^{-1} \ln m$ ,  $0 \leq \vartheta' < \vartheta - 1/2$  and

$$\mathcal{Q}_m(\boldsymbol{\mu}, t; \mathcal{F}) = \left\{ \begin{array}{l} R_m(\rho) = O(m^{\vartheta'}), u_m \geq \frac{\ln \ln m}{\ln m}, 0 < \gamma < 1/2, \\ \pi_{1,m} \geq Cm^{\gamma-0.5}, t \in [0, (\sigma\pi)^{-1} \gamma \ln m] \end{array} \right\}.$$

In each case above,  $C > 0$  can be any constant for which  $\pi_{1,m} \in (0, 1]$  as  $\gamma$  varies in its designated range.

**Corollary 4** provides uniform consistency classes for estimators from Construction I when it is applied to five location-shift families with general scale parameter  $\sigma > 0$ . In particular, if  $\sigma = 1$  and  $q > 3/2$ ,  $\vartheta = q/3$  and  $\vartheta' = 0$  is set in (21), then we recover the uniform consistency class given by Theorems 1.4 and 1.5 of [20].

#### 4. Construction II and Construction III

When the CDFs of  $\{z_i\}_{i=1}^m$  do not have RL type CFs, Construction I in Section 3 cannot be used. In particular, outside location-shift families, the translation-convolution equivalence induced by Fourier transform does not hold and Hoeffding's inequality is no longer applicable. This makes the construction of a proportion estimator using the Strategy much more challenging. So, we will restrict our attention to  $\{z_i\}_{i=1}^m$  whose CDFs do not have RL type CFs but belong to NEFs whose mean and variance are functionally related. Specifically, we show that the Strategy is implementable for discrete NEFs with infinite supports or continuous NEFs with “separable moment functions” (see Definition 4). These include 8 out of the total of 12 NEF-CVFs. The techniques of construction mainly use generating functions and Mellin transform.

##### 4.1. A brief review on natural exponential families

We provide a very brief review on NEF, whose details can be found in [26]. Let  $\beta$  be a positive Radon measure on  $\mathbb{R}$  that is not concentrated on one point. Let  $L(\theta) = \int e^{\theta x} \beta(dx)$  for  $\theta \in \mathbb{R}$  be its Laplace transform and  $\Theta$  be the maximal open set containing  $\theta$  such that  $L(\theta) < \infty$ . Suppose  $\Theta$  is not empty and let  $\kappa(\theta) = \ln L(\theta)$  be the cumulant function of  $\beta$ . Then

$$\mathcal{F} = \{G_\theta : G_\theta(dx) = \exp\{\theta x - \kappa(\theta)\} \beta(dx), \theta \in \Theta\}$$

forms an NEF with respect to the basis  $\beta$ . Note that  $\Theta$  has a non-empty interior if it is not empty and that  $L$  is analytic on the strip  $A_\Theta = \{z \in \mathbb{C} : \Re(z) \in \Theta\}$ .

The NEF  $\mathcal{F}$  can be equivalently characterized by its mean domain and variance function. Specifically, the mean function  $\mu : \Theta \rightarrow U$  with  $U = \mu(\Theta)$  is given by  $\mu(\theta) = \frac{d}{d\theta} \kappa(\theta)$ , and the variance function is  $V(\theta) = \frac{d^2}{d\theta^2} \kappa(\theta)$  and can be parametrized by  $\mu$  as

$$V(\mu) = \int (x - \mu)^2 F_\mu(dx) \text{ for } \mu \in U,$$

where  $\theta = \theta(\mu)$  is the inverse function of  $\mu$  and  $F_\mu = G_{\theta(\mu)}$ . Namely,  $\mathcal{F} = \{F_\mu : \mu \in U\}$ . The pair  $(V, U)$  is called the variance function of  $\mathcal{F}$ , and it characterizes  $\mathcal{F}$ .

For the constructions of proportion estimators for NEFs, we will also use the notation  $K$  but take  $K$  as a function of  $t$  and  $\theta \in \Theta$ . Note that  $K$  depends on  $\theta_0$  but not on any  $\theta \neq \theta_0$ . Further, we will still use the notation  $\psi$  but take it as a function of  $t$  and  $\theta \in \Theta$ . For an NEF,  $\psi$  defined by (1) becomes

$$\psi(t, \theta; \theta_0) = \int K(t, x; \theta_0) dG_\theta(x) \text{ for } G_\theta \in \mathcal{F}.$$

Let  $\theta = (\theta_1, \dots, \theta_m)$ . Then accordingly  $\hat{\varphi}_m(t, \mathbf{z}) = \frac{1}{m} \sum_{i=1}^m \{1 - K(t, z_i; \theta_0)\}$  and  $\varphi_m(t, \theta) = \frac{1}{m} \sum_{i=1}^m \{1 - \psi(t, \theta_i; \theta_0)\}$ .

##### 4.2. Construction II: discrete NEFs with infinite supports

Suppose the basis  $\beta$  for  $\mathcal{F}$  is discrete with support  $\mathbb{N}$ , i.e., there exists a positive sequence  $\{c_k\}_{k \geq 0}$  such that

$$\beta = \sum_{k=0}^{\infty} c_k \delta_k. \quad (22)$$

Then the power series  $H(z) = \sum_{k=0}^{\infty} c_k z^k$  with  $z \in \mathbb{C}$  must have a positive radius of convergence  $R_H$ , and  $H$  is the generating function (GF) of  $\beta$ . Further, if  $\beta$  is a probability measure, then  $(-\infty, 0] \subseteq \Theta$  and  $R_H \geq 1$ , and vice versa. The following approach, which we refer to as “Construction II”, provides the construction for discrete NEFs with support  $\mathbb{N}$ .

**Theorem 4.** Let  $\mathcal{F}$  be the NEF generated by  $\beta$  in (22) and  $\omega$  be admissible. For  $x \in \mathbb{N}$  and  $t \in \mathbb{R}$  set

$$K(t, x; \theta_0) = H(e^{\theta_0}) \int_{[-1, 1]} \frac{(ts)^x \cos\left(\frac{\pi x}{2} - tse^{\theta_0}\right)}{H^{(x)}(0)} \omega(s) ds. \quad (23)$$

Then

$$\psi(t, \theta; \theta_0) = \int K(t, x; \theta_0) dG_\theta(x) = \frac{H(e^{\theta_0})}{H(e^\theta)} \int_{[-1, 1]} \cos\{st(e^\theta - e^{\theta_0})\} \omega(s) ds,$$

$\psi(t, \theta_0; \theta_0) = 1$  for any  $t$ , and  $\lim_{t \rightarrow \infty} \psi(t, \theta; \theta_0) = 0$  for each  $\theta \neq \theta_0$ .

In [Theorem 4](#),  $H^{(k)}(0) = c_k k!$  for  $k \in \mathbb{N}$ . So, for Construction II, if  $\beta$  is known, then we can use  $c_k k!$  instead of  $H^{(k)}(0)$ , whereas if  $H$  is known and  $H^{(k)}$  is easy to compute, we can use  $H^{(k)}(0)$ . This will greatly aid the numerical implementation of Construction II.

#### 4.3. Construction II: some examples

[Theorem 4](#) covers the construction for Abel, Negative Binomial, Poisson, Strict Arcsine, Large Arcsine and Takács families, each of which is an NEF-CVF, has basis  $\beta$  with support  $\mathbb{N}$ , is infinite divisible such that  $H(e^\theta) \neq 0$  for each  $\theta \in \Theta$ , and has non-RL type CFs; see [\[27\]](#) for details on these distributions. However, for each of Abel and Large Arcsine families, the corresponding GF is a composition of two analytic functions, and manually computing  $H(e^{\theta_0})$  in the construction of  $K$  in the statement of [Theorem 4](#) may be cumbersome.

**Example 6.** Poisson family  $\mathcal{P}(\mu)$  with mean  $\mu > 0$ , for which  $F_\mu(\{k\}) = \Pr(X = k) = \frac{\mu^k}{k!} e^{-\mu}$  for  $k \in \mathbb{N}$ . Clearly,

$$\hat{F}_\mu(t) = \exp\{\mu(e^{it} - 1)\} = \exp\{\mu(\cos t - 1)\} \exp(i\mu \sin t).$$

However,  $\{\hat{F}_\mu : \mu > 0\}$  are not of RL type since when  $\mu = \mu_0 + 1$  and  $t > 0$ ,

$$\begin{aligned} \frac{1}{t} \Re \left( \int_{[-t, t]} \frac{\hat{F}_\mu(y)}{\hat{F}_{\mu_0}(y)} dy \right) &= \frac{e^{-1}}{t} \int_{[-t, t]} \exp(\cos y) \cos(\sin y) dy \\ &\geq 2e^{-1} e^{-1} \cos 1 > 0. \end{aligned}$$

The basis is  $\beta = \sum_{k=0}^{\infty} (k!)^{-1} \delta_k$ ,  $L(\theta) = \exp(e^\theta)$  with  $\theta \in \mathbb{R}$ ,  $\mu(\theta) = e^\theta$ ,  $H(z) = e^z$  with  $R_H = \infty$  and  $H^{(k)}(0) = 1$  for all  $k \in \mathbb{N}$ . The Poisson family has been used to model RNA-Seq data [\[10\]](#).

**Example 7.** Negative Binomial family  $\text{NB}(\theta, n)$  with  $\theta < 0$  and  $n \in \mathbb{N}_+$  such that

$$G_\theta(\{k\}) = \Pr(X = k) = \frac{c_k^*}{k!} e^{k\theta} (1 - e^\theta)^n$$

with  $c_k^* = (k + n - 1)! / (n - 1)!$  for  $k \in \mathbb{N}$ . The basis is  $\beta = \sum_{k=0}^{\infty} c_k^* (k!)^{-1} \delta_k$ ,  $L(\theta) = (1 - e^\theta)^{-n}$  with  $\theta < 0$ ,  $\mu(\theta) = ne^\theta (1 - e^\theta)^{-1}$ ,  $H(z) = (1 - z)^{-n}$  with  $R_H = 1$ , and  $H^{(k)}(0) = c_k^*$  for all  $k \in \mathbb{N}$ . The Negative Binomial family has also been used to model RNA-Seq data [\[13,36\]](#).

**Example 8.** Strict Arcsine family. Its VF is  $V(u) = u(1 + u^2)$  and  $\beta = \sum_{n=0}^{\infty} (n!)^{-1} c_n^* (1) \delta_n$ , where  $c_0^*(1) = c_1^*(1) = 1$ ,

$$c_{2n}^*(\sigma) = \prod_{k=0}^{n-1} (\sigma^2 + 4k^2) \quad \text{and} \quad c_{2n+1}^*(\sigma) = \sigma \prod_{k=0}^{n-1} (\sigma^2 + (2k+1)^2) \quad (24)$$

for  $\sigma > 0$  and  $n \in \mathbb{N}_+$ . Further,  $H(z) = \exp(\arcsin z)$  with  $R_H = 1$ . The Strict Arcsine family has been used to model insurance claims [\[23\]](#).

**Example 9.** Large Arcsine family. Its VF is  $V(u) = u(1 + 2u + 2u^2)$  and  $\beta = \sum_{n=0}^{\infty} c_n \delta_n$  with  $c_n = c_n^*(1 + n) / (n + 1)!$  for  $n \in \mathbb{N}$ , where  $c_n^*(\sigma)$  is defined in (24) and for which  $H(z) = \exp\{\arcsin(h(z))\}$  with  $h(z) = \sum_{k=0}^{\infty} c_k z^{k+1}$ . It can be seen that  $R_H$  must be finite; otherwise,  $\lim_{|z_l| \rightarrow \infty} |h(z_l)| = \infty$  for a sequence  $\{z_l : l \geq 1\}$ , and  $H(z_l)$  cannot be expanded into a convergent power series at  $z_l$  for  $l$  sufficiently large.

**Example 10.** Abel family. Its VF  $V(u) = u(1 + u)^2$ ,  $\beta = \sum_{k=0}^{\infty} c_k \delta_k$  with  $c_k = (1 + k)^{k-1} / k!$  for  $k \in \mathbb{N}$ , and  $H(z) = e^{h(z)}$  with  $h(z) = \sum_{k=0}^{\infty} c_k z^{k+1}$  with  $R_H = e^{-1}$ . The Abel family has been used to model birds' migration patterns and other phenomena [\[33\]](#).

**Example 11.** Takács family. Its VF is  $V(u) = u(1 + u)(1 + 2u)$  and  $\beta = \delta_0 + \sum_{k=1}^{\infty} c_k \delta_k$  with  $c_k = (2k)! / (k!(k+1)!)$  for  $k \in \mathbb{N}_+$ , for which  $H(z) = (1 - \sqrt{1 - 4z}) / 2z$  with  $R_H = 4^{-1}$  and  $z = 0$  is a removable singularity of  $H$ .

The above calculations show that the GFs of Negative Binomial, Strict Arcsine, Large Arcsine, Abel and Takács families all have positive and finite radii of convergence whereas that of Poisson family has infinite radius of convergence. This will be very helpful in determining uniform consistency classes for Construction II and its numerical implementation; see [Corollary 5](#).

#### 4.4. Construction III: continuous NEFs with separable moments

In contrast to NEFs with support  $\mathbb{N}$ , we consider non-location-shift NEFs whose members are continuous distributions. Assume  $0 \in \Theta$ , so that  $\beta$  is a probability measure with finite moments of all orders. Let

$$\tilde{c}_n(\theta) = \frac{1}{L(\theta)} \int x^n e^{\theta x} \beta(dx) = \int x^n dG_\theta(x) \quad \text{for } n \in \mathbb{N} \quad (25)$$

be the moment sequence for  $G_\theta \in \mathcal{F}$ . Note that (25) is the Mellin transform of the measure  $G_\theta$ .

**Definition 4.** If there exist two functions  $\zeta, \xi : \Theta \rightarrow \mathbb{R}$  and a sequence  $\{\tilde{a}_n\}_{n \geq 0}$  that satisfy the following:

- $\xi(\theta) \neq \xi(\theta_0)$  whenever  $\theta \neq \theta_0$ ,  $\zeta(\theta) \neq 0$  for all  $\theta \in U$ , and  $\zeta$  does not depend on any  $n \in \mathbb{N}$ ,
- $\tilde{c}_n(\theta) = \xi^n(\theta) \zeta(\theta) \tilde{a}_n$  for each  $n \in \mathbb{N}$  and  $\theta \in \Theta$ ,
- $\Psi(t, \theta) = \sum_{n=0}^{\infty} \frac{t^n \xi^n(\theta)}{\tilde{a}_n n!}$  is absolutely convergent pointwise in  $(t, \theta) \in \mathbb{R} \times \Theta$ ,

then the moment sequence  $\{\tilde{c}_n(\theta)\}_{n \geq 0}$  is called “separable” (at  $\theta_0$ ).

The concept of separable moment sequence is an analogy to the structured integrand used by Construction II, and the condition on  $\Psi(t, \theta)$  is usually satisfied since  $\sum_{n=0}^{\infty} \frac{t^n \xi^n(\theta)}{n!}$  already is convergent pointwise on  $\mathbb{R} \times \Theta$ . The next approach, which we refer to as “Construction III”, is based on Mellin transform of  $G_\theta$  and applies to NEFs with separable moment sequences.

**Theorem 5.** Assume that the NEF  $\mathcal{F}$  has a separable moment sequence  $\{\tilde{c}_n(\theta)\}_{n \geq 0}$  at  $\theta_0$ , and let  $\omega$  be admissible. For  $t, x \in \mathbb{R}$  set

$$K(t, x; \mu_0) = \frac{1}{\zeta(\theta_0)} \int_{[-1,1]} \sum_{n=0}^{\infty} \frac{(-tsx)^n \cos\left\{\frac{\pi}{2}n + ts\xi(\theta_0)\right\}}{\tilde{a}_n n!} \omega(s) ds.$$

Then

$$\psi(t, \mu; \mu_0) = \int K(t, x; \theta_0) dG_\theta(x) = \frac{\zeta(\theta)}{\zeta(\theta_0)} \int_{[-1,1]} \cos[ts\{\xi(\theta_0) - \xi(\theta)\}] \omega(s) ds,$$

$\psi(t, \theta_0; \theta_0) = 1$  for any  $t$  and  $\lim_{t \rightarrow \infty} \psi(t, \theta; \theta_0) = 0$  for each  $\theta \neq \theta_0$ .

Compared to Constructions I and II, Construction III involves the integral of an infinite series and is more complicated. However, it deals with NEFs that have more complicated structures than the former two.

#### 4.5. Construction III: two examples

We provide two examples of Construction III for Exponential and Gamma families, respectively. Note that Gamma family contains Exponential family and central Chi-square family.

**Example 12.** Exponential family  $\text{Exp}(\mu)$  with mean  $\mu > 0$  and basis  $\beta(dx) = e^{-x} \nu(dx)$ , for which  $L(\theta) = (1 - \theta)^{-1}$  and  $\mu(\theta) = 1 - \theta$  for  $\theta < 1$ . Further,

$$\frac{dF_\mu}{d\nu}(x) = f_\mu(x) = \mu e^{-\mu x} 1_{[0, \infty)}(x)$$

and  $\tilde{c}_n(\mu) = (\mu n!) / \mu^n$ . So,  $L(\mu) = \mu^{-1}$ ,  $\xi(\mu) = \mu^{-1}$ ,  $\tilde{a}_n = n!$  and  $\zeta \equiv 1$ . Setting

$$K(t, x; \mu_0) = \int_{[-1,1]} \omega(s) \sum_{n=0}^{\infty} \frac{(-tsx)^n \cos\left(\frac{\pi}{2}n + \frac{ts}{\mu_0}\right)}{(n!)^2} ds$$

gives

$$\psi(t, \mu; \mu_0) = \int K(t, x; \mu_0) dF_\mu(x) = \int_{[-1,1]} \cos\{ts(\mu^{-1} - \mu_0^{-1})\} \omega(s) ds.$$

**Example 13.** Gamma family  $G(\theta, \sigma)$  with basis  $\beta$  such that

$$\frac{d\beta}{d\nu}(x) = \frac{x^{\sigma-1} e^{-x}}{\Gamma(\sigma)} 1_{(0, \infty)}(x) dx \quad \text{with } \sigma > 0,$$

where  $\Gamma$  is the Euler's Gamma function. So,  $L(\theta) = (1 - \theta)^{-\sigma}$ ,

$$\frac{dG_\theta}{dv}(x) = f_\theta(x) = (1 - \theta)^\sigma \frac{e^{\theta x} x^{\sigma-1} e^{-x}}{\Gamma(\sigma)} 1_{(0,\infty)}(x) \quad (26)$$

for  $\theta < 1$ , and  $\mu(\theta) = \sigma/(1 - \theta)$ . Since

$$\tilde{c}_n(\theta) = (1 - \theta)^\sigma \int_0^\infty \frac{e^{-y} y^{n+\sigma-1}}{\Gamma(\sigma)} \frac{dy}{(1 - \theta)^{n+\sigma-1}} = \frac{\Gamma(n + \sigma)}{\Gamma(\sigma)} \frac{1}{(1 - \theta)^n},$$

we see  $\xi(\theta) = (1 - \theta)^{-1}$ ,  $\tilde{a}_n = \Gamma(n + \sigma)/\Gamma(\sigma)$  and  $\zeta \equiv 1$ . Setting

$$K(t, x; \theta_0) = \int_{[-1,1]} \omega(s) \sum_{n=0}^\infty \frac{(-tsx)^n \Gamma(\sigma) \cos\left(\frac{\pi}{2}n + \frac{ts}{1-\theta_0}\right)}{n! \Gamma(\sigma + n)} ds,$$

we obtain

$$\psi(t, \theta; \theta_0) = \int_{[-1,1]} \cos[ts\{(1 - \theta_0)^{-1} - (1 - \theta)^{-1}\}] \omega(s) ds,$$

for which  $\psi(t, \theta_0; \theta_0) = 1$  for all  $t$  and  $\lim_{t \rightarrow \infty} \psi(t, \mu; \theta_0) = 0$  for each  $\theta \neq \theta_0$ .

Recall (10) of Construction I based on Fourier transform, i.e.,

$$\psi(t, \mu; \mu_0) = \int K(t, x; \mu_0) dF_\mu(x) = \int K(t, y + \mu - \mu_0; \mu_0) dF_{\mu_0}(y),$$

where the action of Fourier transform is seen as the translation  $K(t, x; \mu_0) \mapsto K(t, y + \mu - \mu_0; \mu_0)$ . In contrast, the action of Mellin transform is seen via

$$\begin{aligned} \psi(t, \theta; \theta_0) &= \int K(t, x; \theta_0) dG_\theta(x) \\ &= \int_0^\infty K(t, x; \theta_0) (1 - \theta)^\sigma \frac{e^{-(1-\theta)x} x^{\sigma-1}}{\Gamma(\sigma)} dx \end{aligned} \quad (27)$$

$$= \int K\left(t, \frac{y}{1 - \theta}; \theta_0\right) \beta(dy), \quad (28)$$

where from (27) to (28) scaling  $K(t, x; \theta_0) \mapsto K(t, y(1 - \theta)^{-1}; \theta_0)$  is induced. This comparison clearly shows the action of Mellin transform as the multiplication-convolution equivalence, in contrast to the action of Fourier transform as the translation-convolution equivalence.

## 5. Two non-existence results for the Strategy

In this section, we provide two examples for which the Strategy is not implementable. To state them, we introduce

**Definition 5.** The proposition “PropK”: there exists a  $K : \mathbb{R}^2 \rightarrow \mathbb{R}$  such that  $K$  does not depend on any  $\theta \in \Theta_1$  with  $\theta \neq \theta_0$  and that

$$\psi(t, \theta; \theta_0) = \int K(t, x; \theta_0) dG_\theta(x)$$

satisfies  $\lim_{t \rightarrow \infty} \psi(t, \theta_0; \theta_0) = 1$  and  $\lim_{t \rightarrow \infty} \psi(t, \theta; \theta_0) = 0$  for  $\theta \in \Theta_1$  with  $\theta \neq \theta_0$ , where  $\Theta_1$  is a subset of  $\Theta$  that has a non-empty interior and does not contain  $\theta_0$ .

For the following two families, i.e., Inverse Gaussian and Binomial families, PropK does not hold. Note that the former family is discrete but has a finite support, whereas the latter is continuous but does not have a separable moment sequence.

**Example 14.** Inverse Gaussian family  $IG(\theta, \sigma)$  with scale parameter  $\sigma > 0$  and basis

$$\beta(dx) = x^{-3/2} \exp\left(-\frac{\sigma^2}{2x}\right) \frac{\sigma}{\sqrt{2\pi}} 1_{(0,\infty)}(x) \nu(dx),$$

for which  $L(\theta) = \exp(-\sigma\sqrt{-2\theta})$  with  $\theta < 0$ . So,

$$\frac{dG_\theta}{dv}(x) = f_\theta(x) = \frac{x^{-3/2}}{L(\theta)} \exp\left(-\frac{\sigma^2}{2x}\right) \frac{\sigma}{\sqrt{2\pi}} 1_{(0,\infty)}(x).$$



By a change of variables  $y = \frac{\sigma^2}{2x}$ , we obtain

$$\begin{aligned} \int K(t, x; \theta_0) dG_{\theta}(x) &= \frac{1}{L(\theta)} \frac{\sigma}{\sqrt{2\pi}} \int_0^{\infty} K(t, x; \theta_0) x^{-3/2} \exp\left(-\frac{\sigma^2}{2x}\right) dx \\ &= \frac{1}{\sqrt{\pi} L(\theta)} \int_0^{\infty} K\left(t, \frac{\sigma^2}{2y}; \theta_0\right) y^{-1/2} e^{-y} dy. \end{aligned} \quad (29)$$

Since the integral on the right hand side of (29) is not a function of  $\theta$  for  $\theta \neq \theta_0$ , PropK does not hold and  $K$  does not exist. Note that Gaussian family and Inverse Gaussian family are reciprocal pairs, called so by [27]. Inverse Gaussian family has been used to model the shelf life of products [16].

**Example 15.** Binomial family  $\text{Bin}(\theta, n)$  such that

$$\mathbb{P}(X = k) = \binom{n}{k} \theta^k (1 - \theta)^{n-k} \quad \text{for } \theta \in (0, 1).$$

We will show that if PropK holds, then  $K$  must be a function of  $\theta \in \Theta_1$ , which is a contradiction. Assume PropK holds. Then,

$$\psi(t, \theta; \theta_0) = \int K(t, x; \theta_0) dG_{\theta}(x) = (1 - \theta)^n \sum_{k=0}^n a_k(t; \theta_0) \binom{n}{k} \frac{\theta^k}{(1 - \theta)^k},$$

where  $a_x(t; \theta_0) = K(t, x; \theta_0)$  for  $x \in \{0, \dots, n\}$ .

Let  $d_k(t; \theta_0) = a_k(t; \theta_0) \binom{n}{k}$  and  $q(\theta) = \theta/(1 - \theta)$ . Pick  $n + 1$  distinct values  $\varpi_i, i \in \{0, \dots, n\}$  from  $(0, 1)$  such that  $\varpi_0 = \theta_0$  and  $\varpi_i \in \Theta_1$  for  $i \in \{1, \dots, n\}$ . Further, define the  $(n + 1) \times (n + 1)$  Vandermonde matrix  $\mathbf{V}_{n+1}$  whose  $(i, j)$  entry is  $q^{j-1}(\varpi_{i-1})$ , i.e.,  $\mathbf{V}_{n+1} = (q^{j-1}(\varpi_{i-1}))$ ,  $\mathbf{d}_t = (d_0(t; \theta_0), \dots, d_n(t; \theta_0))^{\top}$  and  $\mathbf{b} = ((1 - \theta_0)^{-n}, 0, \dots, 0)^{\top}$ . Then the determinant  $|\mathbf{V}_{n+1}| \neq 0$ , and the properties of  $K$  imply

$$\lim_{t \rightarrow \infty} \mathbf{V}_{n+1} \mathbf{d}_t = \mathbf{b}.$$

However,  $\mathbf{V}_{n+1} : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$  as a bounded linear mapping is a homeomorphism with the bounded inverse  $\mathbf{V}_{n+1}^{-1}$ . So,

$$\lim_{t \rightarrow \infty} \mathbf{d}_t = \lim_{t \rightarrow \infty} \mathbf{V}_{n+1}^{-1} (\mathbf{V}_{n+1} (\mathbf{d}_t)) = \mathbf{V}_{n+1}^{-1} \mathbf{b}.$$

By Cramer's rule, we obtain

$$\lim_{t \rightarrow \infty} d_n(t; \theta_0) = \frac{(-1)^n |\mathbf{V}_n|}{|\mathbf{V}_{n+1}| (1 - \theta_0)^n} = \frac{(-1)^n (1 - \theta_0)^{-n}}{\prod_{k=1}^n (q(\varpi_k) - q(\theta_0))}, \quad (30)$$

where  $\mathbf{V}_n$  is the submatrix of  $\mathbf{V}_{n+1}$  obtained by removing the first row and last column of  $\mathbf{V}_{n+1}$ . But (30) is a contradiction since  $d_n(t; \theta_0)$  does not depend on any  $\theta \in \Theta_1$  for all  $t$ . For the case  $n = 1$ , we easily see from (30) the contradiction

$$\lim_{t \rightarrow \infty} a_1(t; \theta_0) = \frac{-(1 - \theta_0)^{-1}}{q(\theta) - q(\theta_0)} \quad \text{for each } \theta \in \Theta_1.$$

To summarize, PropK does not hold and  $K$  does not exist.

## 6. Construction II: consistency and speed of convergence

Recall  $H^{(k)}(0) = c_k k!$  for  $k \in \mathbb{N}$ . Call the sequence  $\{(c_k k!)^{-1}, k \in \mathbb{N}\}$  the “reciprocal derivative sequence (of  $H$  at 0)”. The following lemma gives bounds on the magnitudes of this sequence for the examples given in Section 4.3. It will help derive concentration inequalities for estimators from Construction II.

**Lemma 2.** Consider the examples given in Section 4.3. Then  $c_k k! \equiv 1$  for Poisson family, whereas for Negative Binomial, Abel and Takács families with a fixed  $n$  and  $\sigma > 0$ ,

$$\frac{1}{c_k k!} \leq \frac{C}{k!} \quad \text{for all } k \in \mathbb{N}. \quad (31)$$

However, for Strict Arcsine and Large Arcsine families, both with a fixed  $\sigma > 0$ , (31) does not hold. On the other hand, for any  $\tilde{r} > 0$  smaller than the radius  $R_H$  of convergence of  $H$ ,

$$\frac{1}{c_k k!} = \frac{1}{H^{(k)}(0)} \geq \frac{1}{H(\tilde{r})} \frac{\tilde{r}^k}{k!} \quad \text{for all } k \in \mathbb{N}. \quad (32)$$

**Lemma 2** shows that, among the six discrete NEFs with support  $\mathbb{N}$  given in Section 4.3, the reciprocal derivative sequence for Poisson family has the largest magnitude, whereas this sequence for Negative Binomial, Abel and Takács families with a fixed  $n$  and  $\sigma > 0$  are all approximately dominated by the “reciprocal factorial sequence”  $\{1/k! : k \in \mathbb{N}\}$ . Further, **Lemma 2** asserts that the reciprocal derivative sequence dominates the “exponential sequence”  $\{H(\tilde{r}) k!/\tilde{r}^k : k \in \mathbb{N}\}$  for any  $\tilde{r} > 0$  smaller than the radius of convergence of  $H$ .

Let  $\eta = e^\theta$  for  $\theta \in \Theta$ ,  $\eta_i = e^{\theta_i}$  for  $i \in \{0, \dots, m\}$  and  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_m)$ . First, we provide upper bounds on the variance and oscillations of  $\hat{\phi}_m(t, \mathbf{z}) - \phi_m(t, \boldsymbol{\theta})$  when  $t$  is positive and sufficiently large.

**Theorem 6.** Let  $\mathcal{F}$  be an NEF generated by  $\beta$  in (22),  $\{z_i\}_{i=1}^m$  be independent with CDFs  $\{G_{\theta_i}\}_{i=1}^m$  belonging to  $\mathcal{F}$ ,  $\lambda$  be a positive constant, and  $t$  be positive and sufficiently large.

(i) Let  $\phi_m(L, \boldsymbol{\theta}) = \min_{1 \leq i \leq m} \{L(\theta_i) \eta_i^{1/4}\}$ . If (31) holds, then

$$\mathbb{V} \{ \hat{\phi}_m(t, \mathbf{z}) - \phi_m(t, \boldsymbol{\theta}) \} \leq V_{\Pi,1}^{(m)} = \frac{C \exp(2t \|\boldsymbol{\eta}\|_\infty^{1/2})}{m \sqrt{t} \phi_m(L, \boldsymbol{\theta})}$$

and

$$\Pr \{ |\hat{\phi}_m(t, \mathbf{z}) - \phi_m(t, \boldsymbol{\theta})| \geq \lambda \} \leq \lambda^{-2} V_{\Pi,1}^{(m)}. \quad (33)$$

(ii) Let  $L_{\min}^{(m)} = \min_{1 \leq i \leq m} L(\theta_i)$ . Then for Poisson family,

$$\mathbb{V} \{ \hat{\phi}_m(t, \mathbf{z}) - \phi_m(t, \boldsymbol{\theta}) \} \leq V_{\Pi,2}^{(m)} = \frac{C \exp(t^2 \|\boldsymbol{\eta}\|_\infty^{1/2})}{m L_{\min}^{(m)}} \quad (34)$$

and

$$\Pr \{ |\hat{\phi}_m(t, \mathbf{z}) - \phi_m(t, \boldsymbol{\theta})| \geq \lambda \} \leq \lambda^{-2} V_{\Pi,2}^{(m)}.$$

We remark that the assertion in **Theorem 6** on Poisson family holds for any NEF with support  $\mathbb{N}$  such that  $c_k k! \leq C$  for all  $k \in \mathbb{N}$ . With **Theorem 6**, we derive the uniform consistency classes and speeds of convergence for the estimators from Construction II.

**Theorem 7.** Let  $\mathcal{F}$  be the NEF generated by  $\beta$  in (22),  $\{z_i\}_{i=1}^m$  be independent with CDFs  $\{G_{\theta_i}\}_{i=1}^m$  belonging to  $\mathcal{F}$ , and  $\rho$  be a finite, positive constant.

(i) If (31) holds, then a uniform consistency class is

$$\mathcal{Q}_{\Pi,1}(\boldsymbol{\theta}, t, \pi_{1,m}; \gamma) = \left\{ \begin{array}{l} \|\boldsymbol{\theta}\|_\infty \leq \rho, \pi_{1,m} \geq m^{(\gamma-1)/2}, t = 2^{-1} \|\boldsymbol{\eta}\|_\infty^{-1/2} \gamma \ln m, \\ \lim_{m \rightarrow \infty} t \min_{i \in I_{1,m}} |\eta_0 - \eta_i| = \infty \end{array} \right\}$$

for any fixed  $\gamma \in (0, 1]$ .

(ii) For Poisson family, a uniform consistency class is

$$\mathcal{Q}_{\Pi,2}(\boldsymbol{\theta}, t, \pi_{1,m}; \gamma) = \left\{ \begin{array}{l} \|\boldsymbol{\theta}\|_\infty \leq \rho, \pi_{1,m} \geq m^{(\gamma'-1)/2}, t = \sqrt{\|\boldsymbol{\eta}\|_\infty^{-1/2} \gamma \ln m}, \\ \lim_{m \rightarrow \infty} t \min_{i \in I_{1,m}} |\eta_0 - \eta_i| = \infty \end{array} \right\}$$

for any fixed  $\gamma \in (0, 1)$  and  $\gamma' > \gamma$ .

In **Theorem 7**, the speed of convergence and uniform consistency class depend on  $\|\boldsymbol{\eta}\|_\infty$  and  $\min_{i \in I_{1,m}} |\eta_0 - \eta_i|$ , whereas they only depend on  $\min_{i \in I_{1,m}} |\mu_i - \mu_0|$  for location-shift families. However, when the GF  $H$  of the basis  $\beta$  has finite radius of convergence, their dependence on  $\|\boldsymbol{\eta}\|_\infty$  can be removed, as justified by:

**Corollary 5.** Let  $\mathcal{F}$  be the NEF generated by  $\beta$  in (22),  $\{z_i\}_{i=1}^m$  be independent with CDFs  $\{G_{\theta_i}\}_{i=1}^m$  belonging to  $\mathcal{F}$ , and  $\rho$  be a finite, positive constant. If (31) holds and  $H$  has a finite radius of convergence  $R_H$ , then for positive and sufficiently large  $t$

$$\mathbb{V} \{ \hat{\phi}_m(t, \mathbf{z}) - \phi_m(t, \boldsymbol{\theta}) \} \leq \frac{C \exp(2t R_H^{1/2})}{m \sqrt{t} \phi_m(L, \boldsymbol{\theta})},$$

and a uniform consistency class is

$$\tilde{\mathcal{Q}}_{\Pi,1}(\boldsymbol{\theta}, t, \pi_{1,m}; \gamma) = \left\{ \begin{array}{l} \|\boldsymbol{\theta}\|_\infty \leq \rho, t = 2^{-1} R_H^{-1/2} \gamma \ln m, \\ \pi_{1,m} \geq m^{(\gamma-1)/2}, \min_{i \in I_{1,m}} |\eta_0 - \eta_i| \geq \frac{\ln \ln m}{\sqrt{2 \ln m}} \end{array} \right\}$$

for any  $\gamma \in (0, 1]$ .

Since  $\sup \Theta \leq \ln R_H$  and  $\|\eta\|_\infty \leq R_H$ , the proof [Corollary 5](#) follows easily from the first claims of [Theorems 6](#) and [7](#) and is omitted. The uniform consistency class  $\tilde{\mathcal{Q}}_{\text{ll},3}(\theta, t, \pi_{1,m}, \gamma)$  is fully data-adaptive and only requires information on  $\min_{i \in I_{1,m}} |\eta_0 - \eta_i|$ , as do Jin's estimator of [\[20\]](#) for Gaussian family and Construction I for location-shift families on  $\min_{i \in I_{1,m}} |\mu_i - \mu_0|$ .

## 7. Construction III: consistency and speed of convergence

We will focus on Gamma family and show that the corresponding estimators are uniformly consistent. Recall  $f_\theta$  defined by [\(26\)](#) for Gamma family. Then  $\hat{f}_\theta(x) = O(x^{\sigma-1})$  as  $x \rightarrow 0+$ , which tends to 0 when  $\sigma > 1$ . The next result provides upper bounds on the variance and oscillations of  $\hat{\varphi}_m(t, \mathbf{z}) - \varphi_m(t, \theta)$  when  $t$  is positive and large.

**Theorem 8.** Consider Gamma family such that  $\{z_j\}_{j=1}^m$  are independent with parameters  $\{(\theta_i, \sigma)\}_{i=1}^m$  and a fixed  $\sigma > 0$ . Assume  $t$  is positive and sufficiently large and set  $u_{3,m} = \min_{1 \leq i \leq m} \{1 - \theta_i\}$ . Then

$$\mathbb{V} \{ \hat{\varphi}_m(t, \mathbf{z}) - \varphi_m(t, \theta) \} \leq V_{\text{III}} = \frac{C}{m^2} \exp \left( \frac{4t}{u_{3,m}} \right) \sum_{i=1}^m \left( \frac{t}{1 - \theta_i} \right)^{3/4 - \sigma}$$

and

$$\Pr \{ | \hat{\varphi}_m(t, \mathbf{z}) - \varphi_m(t, \theta) | \geq \lambda \} \leq \lambda^{-2} V_{\text{III}}. \quad (35)$$

Recall  $\xi(\theta) = (1 - \theta)^{-1}$ , such that  $u_{3,m} = \min_{1 \leq i \leq m} \xi^{-1}(\theta_i)$ . Using [Theorem 8](#), we show the uniform consistency and speed of convergence of the estimator for Gamma family.

**Theorem 9.** Consider Gamma family such that  $\{z_j\}_{j=1}^m$  are independent with parameters  $\{(\theta_i, \sigma)\}_{i=1}^m$  and a fixed  $\sigma > 0$ . Let  $\rho > 0$  be a finite constant. If  $\sigma > 3/4$ , then a uniform consistency class

$$\mathcal{Q}_{\text{III}}(\theta, t, \pi_{1,m}; \gamma) = \left\{ \begin{array}{l} \|\theta\|_\infty \leq \rho, t = 4^{-1} \gamma u_{3,m} \ln m, \lim_{m \rightarrow \infty} u_{3,m} \ln m = \infty, \\ \pi_{1,m} \geq m^{(\gamma-1)/2}, \lim_{m \rightarrow \infty} t \min_{i \in I_{1,m}} |\xi(\theta_0) - \xi(\theta_i)| = \infty \end{array} \right\}$$

for any fixed  $\gamma \in (0, 1]$ . On the other hand, if  $\sigma \leq 3/4$ , then

$$\mathcal{Q}_{\text{III}}(\theta, t, \pi_{1,m}; \gamma) = \left\{ \begin{array}{l} \|\theta\|_\infty \leq \rho, t = 4^{-1} \gamma u_{3,m} \ln m, \pi_{1,m} \geq m^{(\gamma'-1)/2}, \\ \lim_{m \rightarrow \infty} t \min_{i \in I_{1,m}} |\xi(\theta_0) - \xi(\theta_i)| = \infty \end{array} \right\}$$

for any fixed  $\gamma \in (0, 1)$  and  $\gamma' > \gamma$ .

In [Theorem 9](#), the speed of convergence and uniform consistency class depend on  $u_{3,m} = \min_{1 \leq i \leq m} \{1 - \theta_i\}$  and  $\xi_{3,m} = \min_{i \in I_{1,m}} |\xi(\theta_0) - \xi(\theta_i)|$ . Since  $\theta < 1$  for Gamma family,  $u_{3,m}$  measures how close a  $G_{\theta_i}$  is to the singularity where a Gamma density is undefined, and it is sensible to often assume  $\liminf_{m \rightarrow \infty} u_{3,m} > 0$ . On the other hand,  $\sigma \xi(\theta) = \mu(\theta)$  for all  $\theta \in \Theta$  for Gamma family. So,  $\xi_{3,m}$  measures the minimal difference between the means of  $G_{\theta_i}$ ,  $i \in I_{1,m}$  and  $G_{\theta_0}$ , and  $\xi_{3,m}$  cannot be too small relative to  $t$  as  $t \rightarrow \infty$  in order for the estimator to achieve consistency.

## 8. Simulation studies

We present a simulation study on  $\hat{\varphi}_m(t_m, \mathbf{z})$ , with comparison to the “MR” estimator of [\[32\]](#) and the “hybrid estimator” induced by “Jin’s estimator” of [\[20\]](#). Since the MR estimator is only applicable to p-values that have continuous distributions, its performance provides information on its robustness when it is applied to discrete p-values such as those induced by Poisson and Negative Binomial distributions (to be considered hereunder). Note that when  $X_0$  is a realization of a Poisson or Negative Binomial random variable with CDF  $F_0^*$ , we set its p-value as  $F_0^*(X_0)$ .

### 8.1. Simulation design

For  $a < b$ , let  $U(a, b)$  be the uniform random variable or the uniform distribution on the closed interval  $[a, b]$ . We consider 7 values for  $m$ , i.e.,  $m = 10^3, 5 \times 10^3, 10^4, 5 \times 10^4, 10^5, 5 \times 10^5$  or  $10^6$ , and 4 sparsity levels for  $\pi_{1,m}$ , i.e., the dense regime  $\pi_{1,m} = 0.2$ , moderately sparse regime  $\pi_{1,m} = m^{-0.2}$ , critically sparse regime  $\pi_{1,m} = m^{-0.5}$  and very sparse regime  $\pi_{1,m} = m^{-0.7}$ , where we recall  $\pi_{1,m} = 1 - m_0 m^{-1}$  and  $m_0 = |\{i \in \{1, \dots, m\} : \mu_i = \mu_0\}|$ . Further, we consider 5 distribution families  $\mathcal{F}$ , i.e., Laplace, Cauchy, Poisson, Negative Binomial and central Chi-square families, set as follows:

- For Laplace and Cauchy families,  $\sigma = 1$  and  $\mu_0 = 0$  is set, and the nonzero  $\mu_i$ 's are generated independently such that their absolute values  $|\mu_i|$  are from  $U(0.75, 5)$  but each  $\mu_i$  has probability 0.5 to be negative or positive.

- For Negative Binomial family,  $n = 5$  and  $\theta_0 = -4.5$ ; for Poisson family,  $\theta_0 = 0.08$ ; for Gamma family,  $\theta_0 = 0.5$  and  $\sigma = 6$ . Note that for Gamma family, under the null hypothesis the corresponding distribution is a central Chi-square distribution with  $2^{-1}\sigma$  degrees of freedom. For all three families,  $\theta_i = \theta_0 \rho_i$  for  $i \in \{m_0 + 1, \dots, m_0 + \lfloor 2^{-1}m\pi_m \rfloor\}$  and  $\theta_i = \theta_0 \rho_i^{-1}$  for  $i \in \{m_0 + \lfloor 2^{-1}m\pi_m \rfloor + 1, \dots, m\}$ , where  $\{\rho_i\}_{i=m_0+1}^m$  are independently generated from  $U(10, 13)$  for Poisson family, from  $U(8, 15)$  for Negative Binomial family, and from  $U(1.2, 1.5)$  for Gamma family.

Each triple  $(m, \pi_m, \mathcal{F})$  gives an experiment, and there are 140 experiments. Each experiment is repeated independently 250 times so that summary statistics can be obtained. For an estimator  $\hat{\pi}_{1,m}$  of  $\pi_{1,m}$  for  $m \geq 1$  for each experiment, the mean and standard deviation of  $\tilde{\delta}_m = \hat{\pi}_{1,m} \pi_{1,m}^{-1} - 1$  is estimated from the 250 realizations.

Details on the implementations of the estimators to be compared are given below. For the estimator  $\hat{\varphi}_m(t_m, \mathbf{z})$ , the sequence  $\{t_m\}_{m \geq 1}$  is set by Corollary 3 as  $t_m = \ln m$  for Laplace family and as  $t_m = \sqrt{\ln m}$  for Cauchy family, by Theorem 7 as  $t_m = \|\eta\|_\infty^{-1/4} \sqrt{\ln m}$  for Poisson family, by Corollary 5 as  $t_m = 2^{-1} \ln m$  for Negative Binomial family, and by Theorem 9 to be  $t_m = 4^{-1} \ln m \min_{1 \leq i \leq m} \{1 - \theta_i\}$  for central Chi-square family, where the tuning parameter  $\gamma$  for  $\hat{\varphi}_m(t_m, \mathbf{z})$  has been set to be the supremum of its feasible range to enable the estimator to achieve a fast speed of convergence. The averaging function  $\omega$  is chosen to be the triangular density on  $[-1, 1]$ . Further, from Section 3.1, Theorem 4 and Section 4.5, we see the following: for Cauchy family with  $\mu_0 = 0$  and  $\sigma = 1$ ,

$$K(t, x; 0) = \int_{[-1, 1]} \exp(|t|) \omega(s) \cos(tsx) ds; \quad (36)$$

for Laplace family with  $\mu_0 = 0$  and  $\sigma = 1$ ,

$$K(t, x; 0) = \int_{[-1, 1]} (1 + t^2) \omega(s) \cos(tsx) ds; \quad (37)$$

for Poisson family and Negative Binomial family

$$K(t, x; \theta_0) = H(e^{\theta_0}) \int_{[-1, 1]} \frac{(ts)^x \cos\left(\frac{\pi x}{2} - tse^{\theta_0}\right)}{H^{(x)}(0)} \omega(s) ds; \quad (38)$$

for central Chi-square family

$$K(t, x; \theta_0) = \int_{[-1, 1]} \omega(s) \sum_{n=0}^{\infty} \frac{(-tsx)^n \Gamma(\sigma) \cos\left(\frac{\pi}{2}n + \frac{ts}{1-\theta_0}\right)}{n! \Gamma(\sigma + n)} ds. \quad (39)$$

For the integrals in (36), (37), (38) and (39), each integral is approximated by a Riemann sum for which the interval  $[-1, 1]$  is partitioned into 400 equal subintervals, for which each integrand is evaluated at the end points of these subintervals. Further, for the integral in (39), the power series in the integrand is replaced by the partial sum of its first 21 terms, i.e., it is truncated at  $n = 20$ . The MR estimator (defined for continuous p-values) is implemented as follows: let the ascendingly ordered p-values be  $p_{(1)} < p_{(2)} < \dots < p_{(m)}$  for  $m > 4$ , set  $b_m^* = m^{-1/2} \sqrt{2 \ln \ln m}$  and define

$$q_i^* = (1 - p_{(i)})^{-1} \left\{ im^{-1} - p_{(i)} - b_m^* \sqrt{p_{(i)} (1 - p_{(i)})} \right\};$$

then  $\hat{\pi}_{1,m}^{mr} = \min\{1, \max\{0, \max_{2 \leq i \leq m-2} q_i^*\}\}$  is the MR estimator. Note that the MR estimator implicitly assumes that the probability for any tie between the p-values is zero and that when it is applied to discrete p-values, it is OK to allow for such ties. The hybrid estimator is implemented as follows: first, each  $z_i$  is transformed into  $\tilde{z}_i = \Phi^{-1}(F_{i0}(z_i))$ , where  $F_{i0}$  is the CDF of  $z_i$  under the null hypothesis and  $\Phi^{-1}$  the inverse of the CDF of the standard Normal random variable; secondly, Jin's estimator of [20] for Gaussian family is applied to  $\{\tilde{z}_i\}_{i=1}^m$ , for which the integral in (8) is approximated by a Riemann sum based on partitioning  $[-1, 1]$  into subintervals of equal length 0.01 (the default setting in [20]),  $\omega$  is set as the triangular density on  $[-1, 1]$ , and  $\gamma = 0.5$  is set in  $t_m = \sqrt{2\gamma \ln m}$  to provide the fastest possible speed of convergence.

## 8.2. Simulation results

For an estimator  $\hat{\pi}_{1,m}$  of  $\pi_{1,m}$ , we will measure its stability by the standard deviation  $\sigma_m^*$  of  $\tilde{\delta}_m$  and its accuracy by the mean  $\mu_m^*$  of  $\tilde{\delta}_m$ . Among two estimators for a fixed  $m$ , the one that has both smaller  $\sigma_m^*$  and  $|\mu_m^*|$  will be considered better. The supplementary material contains boxplots that summarize the performances of the three estimators under investigation.

The following five observations have been made from the comparison between the proposed estimator and the MR estimator: (i) For Laplace and Cauchy families, the new estimator is very accurate and much better than the MR estimator. In the dense and moderately sparse regimes, there is very strong evidence on the consistency of the new estimator since  $\tilde{\delta}_m$  displays a strong trend to converge to 0. These may hold true when the new estimator is applied to other location-shift families. (ii) For Poisson and Negative Binomial families, the new estimator is accurate and much better than the MR

estimator. In the dense and moderately sparse regimes, there is strong evidence on the convergence of  $\tilde{\delta}_m$  (even though not necessarily to 0) when the new estimator is applied to Negative Binomial family, whereas there is no strong evidence of convergence of  $\tilde{\delta}_m$  when it is applied to Poisson family. In other words, we have not observed strong evidence on the consistency of the new estimator. This may be a consequence of non-adaptively choosing  $t_m$  for the estimator, and is worth further investigation. In contrast, the MR estimator is almost always zero, i.e., it is rarely able to detect the existence of false null hypotheses. (iii) For Gamma family, the new estimator is more accurate than the MR estimator, and it often severely underestimates  $\pi_{1,m}$ . When the new estimator is applied to the dense and moderately sparse regimes, there is strong evidence on the convergence of  $\tilde{\delta}_m$  but there is no strong evidence on the consistency of the new estimator. This may be due to truncating the power series in Construction III when implementing the estimator for Gamma family and non-adaptively choosing  $t_m$  for the estimator, and requires further investigation. In contrast, the MR estimator is almost always zero, often failing to detect the existence of false null hypotheses. Such an interesting behaviour for the MR estimator has not been reported before. (iv) The MR estimator, if not being 0 for all almost all repetitions of an experiment, is less stable than the new estimator and can be much so when  $m$  is large. In the critically sparse and very sparse regimes, the new estimator does not seem to be consistent and its  $\tilde{\delta}_m$  does not display a trend of convergence as  $m$  increases. Similarly, in these regimes, the MR estimator does not seem to be consistent unless it is identically zero. However, this does not contradict the theory for the MR estimator since its consistency requires that p-values under the alternative hypothesis be identically distributed and they are not so in the simulation study here. (v) In the dense and moderately sparse regimes, the new estimator usually underestimates  $\pi_{1,m}$ , i.e., the estimated proportion of true null hypotheses,  $\hat{\pi}_{0,m}$ , induced by the estimator is usually conservative. This is appealing in that a one-step adaptive FDR procedure that employs  $\hat{\pi}_{0,m}$  is usually conservative.

On the other hand, the following three observations have been made from the comparison between the proposed estimator and the hybrid estimator: (i) when  $z_i$ 's have Laplace or Cauchy distributions, the proposed estimator is (much) more accurate but a bit less stable than the hybrid estimator in the dense and moderately sparse regimes, whereas they have competitive performances in the critically and very sparse regimes; (ii) when  $z_i$ 's have central Chi-square distributions, the hybrid estimator is more accurate but a bit less stable than the proposed estimator in the dense and moderately sparse regimes, whereas they have competitive performances in the critically and very sparse regimes. However, for this scenario, the proposed estimator is implemented by truncating the power series in the integral in (39) at the 20th term, and its performance can be improved by better approximating the power series; (iii) when  $z_i$ 's have Poisson or Negative Binomial distributions, the proposed estimator is much more accurate than and as stable as the hybrid estimator across all sparsity regimes. This is reasonable since when  $z_i$ 's have discrete CDFs, the assumptions on the continuity and Normality of the transformed random variables  $\tilde{z}_i$ 's are violated, and Jin's method is not applicable. We remark that when  $z_i$ 's have central Chi-square distributions such that  $\theta_0 = 0.05$  and  $\sigma = 9$  are set in Section 8.1, the proposed estimator is more accurate and stable than the hybrid estimator in the dense and moderately sparse regimes, whereas they have competitive performances in the critically and very sparse regimes. In summary, the hybrid estimator is not able to serve as the universal estimator, and tailored ones such as the proposed are needed for specific scenarios.

## 9. Discussion

We have demonstrated that solutions of Lebesgue–Stieltjes integral equations can serve as a universal construction for proportions estimators, provided proportion estimators for random variables with three types of distributions, and justified under independence the uniform consistency and speeds of convergence of the estimators. For a proposed estimator to achieve uniform consistency, the tuning parameter that determines its intrinsic speed of convergence needs to be determined adaptively based on data. On the other hand, in applications we usually have information from domain scientists on a lower bound on the proportion of false null hypotheses and on the minimal effect size. Further, for each estimator from each construction, an upper bound on its variance with respect to the oracle has been provided, and the minimal effect size to ensure its uniform consistency can converge to zero quite fast. So, inspired by the work of [20], we can adaptively determine the tuning parameter so that the intrinsic speed forces the variance upper bound to converge to zero at certain rate and to be of smaller order than the lower bound on the proportion, thus achieving uniform consistency adaptively. Specifically, this can be done for estimators from Construction I, for estimators from Construction II where generating functions have finite radii of convergence, and for other estimators of Construction II and those from Construction III after estimating the supremum norm of the parameter vector or the infimum of a known transform of the vector itself. In an accompanying article, we will deal with this estimation problem, discuss how truncating the power series in Construction III affects the accuracy of the induced estimators, and report via extensive simulation studies the adaptive, non-asymptotic performances of the proposed estimators.

Our work induces three topics that are worthy of future investigations. Firstly, we have only considered estimating the proportion of parameters that are unequal to a fixed value, i.e., the proportion induced by the functional that maps a parameter to a fixed value. It would be interesting to construct uniformly consistent estimators of proportions induced by other functionals. Further, we have only considered independent random variables. Extending the consistency results provided here to dependent case will greatly enlarge the scope of applications of the estimators, as did by [7] and [21] to Jin's estimator of [20] for Gaussian family and Gaussian mixtures. Moreover, Construction I, II and III are applicable to random variables whose distributions are from different sub-families of the same type of distributions, and results

on the uniform consistency of the corresponding proportion estimators can be extended to this case. Finally, following the principles in Section 3 of [20], Construction I, II and III can possibly be applied to consistently estimate the mixing proportions for two-component mixture models at least one of whose components follows a distribution discussed in this work.

Secondly, we have only been able to construct proportion estimators for three types of distributions, and provide Gamma family as an example for Construction III. It is worthwhile to explore other settings for which solutions of Lebesgue–Stieltjes integro-differential equations exist, can be analytically expressed, and serve as consistent proportion estimators. Further, we have not studied optimal properties of the proposed estimators, and with regard to this the techniques of [4] and [6] may be useful.

Thirdly, we have introduced the concept of “the family of distributions with Riemann–Lebesgue type characteristic functions (RL type CFs)” (see Definition 1) for which

$$\left\{t \in \mathbb{R} : \hat{F}_{\mu_0}(t) = 0\right\} = \emptyset, \quad (40)$$

and shown that it contains several location-shift families. The requirement (40) precludes the characteristic function  $\hat{F}_{\mu_0}$  to have any real zeros. We are aware that Poisson family is infinitely divisible but not a location-shift family and does not have RL type CFs. However, it is unclear to us the relationships (with respect to set inclusion) between infinitely divisible distributions, location-shift families and distributions with RL type CFs. So, a better understanding of such relationships will contribute both to the theory of probability distributions and finding examples different than those given here that Constructions I, II and III apply to.

## Acknowledgments

Part of the research was funded by the New Faculty Seed Grant provided by Washington State University, USA. I would like to thank the Editor and Associate Editor for handling my submission and the reviewers for their helpful comments. I am very grateful to Gérard Letac for his guidance on my research involving natural exponential families and constant encouragements, Jiashun Jin for providing the technical report [19], warm encouragements and comments on the presentation of an earlier version of the manuscript, Mark D. Ward for comments on exponential generating functions, Kevin Vixie, Hong-Ming Yin and Charles N. Moore for discussions on solutions of integro-differential equations, Sheng-Chi Liu for discussions on solutions of algebraic equations, and Ovidiu Costin and Sergey Lapin for help with access to two papers.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jmva.2019.06.003>.

## References

- [1] P. Auer, R. Doerge, Statistical design and analysis of RNA-Seq data, *Genetics* (2010) 405–416.
- [2] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, *Ann. Statist.* 29 (2001) 1165–1188.
- [3] G. Blanchard, E. Roquain, Adaptive false discovery rate control under independence and dependence, *J. Mach. Learn. Res.* 10 (2009) 2837–2871.
- [4] T.T. Cai, J. Jin, Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing, *Ann. Statist.* 38 (2010) 100–145.
- [5] T.T. Cai, W. Sun, Simultaneous testing of grouped hypotheses: Finding needles in multiple haystacks, *J. Amer. Statist. Assoc.* 104 (2009) 1467–1481.
- [6] A. Carpentier, N. Verzelen, Adaptive estimation of the sparsity in the gaussian vector model, [arXiv:1703.00167](https://arxiv.org/abs/1703.00167), 2017.
- [7] X. Chen, Consistent FDR estimation for adaptive multiple testing Normal means under principal correlation structure, [arXiv:1410.4275v4](https://arxiv.org/abs/1410.4275v4), 2018.
- [8] X. Chen, R.W. Doerge, A weighted FDR procedure under discrete and heterogeneous null distributions, [arXiv:1502.00973v4](https://arxiv.org/abs/1502.00973v4), 2017.
- [9] X. Chen, R.W. Doerge, J.F. Heyse, Multiple testing with discrete data: proportion of true null hypotheses and two adaptive FDR procedures, *Biometrical J.* 60 (2018) 761–779.
- [10] X. Chen, J.D. Storey, Consistent estimation of low-dimensional latent structure in high-dimensional data, [arXiv:1510.03497](https://arxiv.org/abs/1510.03497), 2015.
- [11] A. Cohen, Y. Rabinovich, A. Schuster, H. Shachnai, Optimal bounds on tail probabilities: A study of an approach, in: *Optimal Bounds on Tail Probabilities: A Study of an Approach*, Springer US, Boston, MA, 1999, pp. 1–24.
- [12] O. Costin, N. Falkner, J.D. McNeal, Some generalizations of the Riemann–Lebesgue lemma, *Am. Math. Mon.* 123 (2016) 387–391.
- [13] Y. Di, D.W. Schafer, J.S. Cumbie, J.H. Chang, The NBP negative binomial model for assessing differential gene expression from RNA-Seq, *Stat. Appl. Genet. Mol. Biol.* 10 (2011).
- [14] B. Efron, R. Tibshirani, J.D. Storey, V. Tusher, Empirical bayes analysis of a microarray experiment, *J. Amer. Statist. Assoc.* 96 (2001) 1151–1160.
- [15] M.J. Fischer, *Generalized Hyperbolic Secant Distributions*, Springer, 2014.
- [16] J.L. Folks, R.S. Chhikara, The inverse Gaussian distribution and its statistical application—a review, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 40 (1978) 263–289.
- [17] C. Genovese, L. Wasserman, A stochastic process approach to false discovery control, *Ann. Statist.* 32 (2004) 1035–1061.
- [18] P.B. Gilbert, A modified false discovery rate multiple-comparisons procedure for discrete data, applied to human immunodeficiency virus genetics, *J. R. Stat. Soc. Ser. C Appl. Stat.* 54 (2005) 143–158.
- [19] J. Jin, Proportions of nonzero normal means: universal oracle equivalence and uniformly consistent estimations, Technical report. Department of statistics, Purdue University, West Lafayette, 2006.
- [20] J. Jin, Proportion of non-zero normal means: universal oracle equivalences and uniformly consistent estimators, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 70 (2008) 461–493.



- [21] J. Jin, T.T. Cai, Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons, *J. Amer. Statist. Assoc.* 102 (2007) 495–506.
- [22] J. Jin, J. Peng, P. Wang, A generalized fourier approach to estimating the null parameters and proportion of nonnull effects in large-scale multiple testing, *J. Stat. Res.* 44 (2010) 103–107.
- [23] C.C. Kokonendji, M. Khoudar, On strict arcsine distribution, *Comm. Statist. Theory Methods* 33 (2004) 993–1006.
- [24] P.R. Kumar, S. Bodhisattva, Estimation of a two-component mixture model with applications to multiple testing, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 78 (2016) 869–893.
- [25] M. Langaas, L.B. Henery, E. Ferkingstad, Estimating the proportion of true null hypotheses, with application to dna microarray data, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67 (2005) 555–572.
- [26] G. Letac, Lectures on natural exponential families and their variance functions, in: *Monografias de matemática*, vol. 50, IMPA, Rio de Janeiro, 1992.
- [27] G. Letac, M. Mora, Natural real exponential families with cubic variance functions, *Ann. Statist.* 18 (1990) 1–37.
- [28] Y. Liu, S.K. Sarkar, Z. Zhao, A new approach to multiple testing of grouped hypotheses, *J. Statist. Plann. Inference* 179 (2016) 1–14.
- [29] E. Lukacs, *Characteristic Functions*, Hafner Publishing Company, 1970.
- [30] S. Luo, Z. Zhang, Estimating the first zero of a characteristic function, *C. R. Math.* 338 (2004) 203–206.
- [31] D.V. Mehrotra, J.F. Heyse, Use of the false discovery rate for evaluating clinical safety data, *Stat. Methods Med. Res.* 13 (2004) 227–238.
- [32] N. Meinshausen, J. Rice, Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses, *Ann. Statist.* 34 (2006) 373–393.
- [33] S.B. Nandi, K.K. Das, A family of the abel series distributions, *Sankhya B* 56 (1994) 147–164.
- [34] J. Pitman, M. Yor, Infinitely divisible laws associated with hyperbolic functions, *Canad. J. Math.* 55 (2003) 292–330.
- [35] A. Ploner, S. Calza, A. Gusnanto, Y. Pawitan, Multidimensional local false discovery rate for microarray studies, *Bioinformatics* 22 (2006) 556–565.
- [36] M.D. Robinson, G.K. Smyth, Small-sample estimation of negative binomial dispersion, with applications to SAGE data, *Biostatistics* 9 (2008) 321–332.
- [37] S.K. Sarkar, False discovery and false nondiscovery rates in single-step multiple testing procedures, *Ann. Statist.* 34 (2006) 394–415.
- [38] J.D. Storey, A direct approach to false discovery rates, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64 (2002) 479–498.
- [39] J. Storey, The positive false discovery rate: a Bayesian interpretation and the q-value, *Ann. Statist.* 3 (2003) 2013–2035.
- [40] J.D. Storey, The optimal discovery procedure: a new approach to simultaneous significance testing, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 69 (2007) 347–368.
- [41] J.D. Storey, J.E. Taylor, D. Siegmund, Strong control, conservative point estimation in simultaneous conservative consistency of false discovery rates: a unified approach, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 66 (2004) 187–205.
- [42] J.W.H. Swanepoel, The limiting behavior of a modified maximal symmetric 2s-spacing with applications, *Ann. Statist.* 27 (1999) 24–35.
- [43] D.C. Vaughan, The generalized secant hyperbolic distribution and its properties, *Comm. Statist. Theory Methods* 31 (2002) 219–238.