Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Fisher-Pitman Permutation Tests Based on Nonparametric Poisson Mixtures with Application to Single Cell Genomics

Zhen Miao, Weihao Kong, Ramya Korlakai Vinayak, Wei Sun & Fang Han

To cite this article: Zhen Miao, Weihao Kong, Ramya Korlakai Vinayak, Wei Sun & Fang Han (2024) Fisher-Pitman Permutation Tests Based on Nonparametric Poisson Mixtures with Application to Single Cell Genomics, Journal of the American Statistical Association, 119:545, 394-406, DOI: 10.1080/01621459.2022.2120401

To link to this article: https://doi.org/10.1080/01621459.2022.2120401

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2024, VOL. 119, NO. 545, 394–406: Theory and Methods
https://doi.org/10.1080/01621459.2022.2120401

Fisher-Pitman Permutation Tests Based on Nonparametric Poisson Mixtures with Application to Single Cell Genomics

Zhen Miao[a], Weihao Kong[b], Ramya Korlakai Vinayak[c], Wei Sun[d], and Fang Han[a]

[a]Department of Statistics, University of Washington, Seattle, WA; [b]Google Inc, Mountain View, CA; [c]Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI; [d]Public Health Science Division, Fred Hutchinson Cancer Research Center, Seattle, WA

ABSTRACT

This article investigates the theoretical and empirical performance of Fisher-Pitman-type permutation tests for assessing the equality of unknown Poisson mixture distributions. Building on nonparametric maximum likelihood estimators (NPMLEs) of the mixing distribution, these tests are theoretically shown to be able to adapt to complicated unspecified structures of count data and also consistent against their corresponding ANOVA-type alternatives; the latter is a result in parallel to classic claims made by Robinson. The studied methods are then applied to a single-cell RNA-seq data obtained from different cell types from brain samples of autism subjects and healthy controls; empirically, they unveil genes that are differentially expressed between autism and control subjects yet are missed using common tests. For justifying their use, rate optimality of NPMLEs is also established in settings similar to nonparametric Gaussian (Wu and Yang) and binomial mixtures (Tian, Kong, and Valiant; Vinayak et al.). Supplementary materials for this article are available online.

ARTICLE HISTORY
Received August 2021
Accepted August 2022

KEYWORDS
Fisher-Pitman permutation tests; Minimax risk; Nonparametric MLE; Nonparametric Poisson mixture; Single-cell genomics

1. Introduction

Considering an experiment with multiple samples drawn from multiple populations, distinguishing possible difference among them in one or more dimensions is a fundamental statistical task. In the classical test of the null hypothesis of no mean differences, one-way analysis of variance (ANOVA, see Fisher 1925) F-test is perhaps the most commonly used tool, and is the uniformly most powerful invariant one under additional normal assumption, see Scheffé (1959, p. 50).

Despite its popularity, one-way ANOVA has its competing alternatives. In the context of randomized experiments, Fisher (Fisher 1935) initialized an ingenious permutation approach as an alternative to performing ANOVA F-test. This idea was later developed further by Pitman (Pitman 1938). The resulting procedures, often termed the Fisher-Pitman permutation tests in literature, achieve the appealing property of being exactly distribution-free and have been suggested in various contexts as, for example, when the distributional assumptions of F-tests no longer hold (Marascuilo and McSweeney 1977; Still and White 1981; Berry and Mielke 1983). Robustness properties have been further studied empirically (Boik 1987) and theoretically (Chung and Romano 2013); power analyses were also performed in Hoeffding (1952) and Robinson (1973).

Although being originally defined in Euclidean spaces, it is by now well understood that the ANOVA F-tests and especially their permutation-type alternatives are able to adapt to an arbitrary metric space. This is via the approach of "interpoint" distance functions (Mielke Jr., Berry, and Johnson 1976; Mielke Jr. 1984) that uses an alternative representation of the F statistic as a function of between- and within-group pairwise distances. Thus, through replacing the original Euclidean distance by any properly defined distance function, the idea of Fisher-Pitman permutation tests is now implementable in many complicated metric spaces beyond the Euclidean (Anderson 2001; Mielke and Berry 2007; Petersen and Müller 2019).

Our study of Fisher-Pitman-type permutation tests stems from the analysis of single-cell RNA-seq (scRNA-seq) data, and particularly, a framework that was recently promoted in Sarkar and Stephens (2021). There, the authors described how a separation of measurement and expression models is able to clarify confusion in modeling scRNA-seq data, and accordingly advocated using the terminology of Poisson mixtures to unify many existing models (see Table 1 in Sarkar and Stephens 2021). In detail, thinking about $X_{ij}^{(k)}$ to be the absolute expression of a specific gene in cell $i \in [N_{jk}] := \{1, 2, \ldots, N_{jk}\}$ of subject $j \in [n_k]$ of population $k \in [K]$, we are interested in studying the following model of $X_{ij}^{(k)}$ that is a slight simplification to Sarkar and Stephens's Equation (1):

$$X_{ij}^{(k)} \mid \lambda_{ij}^{(k)} \sim \text{Poisson}\big(r_{ij}^{(k)}\lambda_{ij}^{(k)}\big); \qquad \text{(measurement model)} \quad (1.1)$$

$$\lambda_{ij}^{(k)} \sim Q_j^{(k)}. \qquad \text{(expression model)} \quad (1.2)$$

CONTACT Fang Han fanghan@uw.edu Department of Statistics, University of Washington, Seattle, WA.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

Here $r_{ij}^{(k)} > 0$ adjusts the cell "read depth" (see Zhang, Ntranos, and Tse 2020, p. 1), often set to be the scaled cell's total reads across all genes (Sarkar and Stephens 2021, eq. (1)), and in this article is assumed to be known; $Q_j^{(k)}$ is a properly defined distribution that describes the "expression level" of the gene in population k and is assumed to have a compact support on the nonnegative real line. Adopting the statistical terminology, for each $k \in [K]$ and $j \in [n_k]$, $\{X_{ij}^{(k)}, i = 1, \ldots, N_{jk}\}$ then independently follow Poisson mixture distributions of point mass functions (PMFs)

$$h_{ij}^{(k)}(x) := \int_0^\infty e^{-\lambda r_{ij}^{(k)}} \frac{\{\lambda r_{ij}^{(k)}\}^x}{x!} \mathrm{d}Q_j^{(k)}(\lambda), \quad x = 0, 1, 2, \ldots$$

and a mixing distribution $Q_j^{(k)}$ that has to be characterized by a nonparametric model; see Sarkar and Stephens (2021, sec. "Modeling scRNA-seq data") for a discussion of why a nonparametric model of $Q_j^{(k)}$ is preferred in single-cell genomics, though Sarkar and Stephens (2021) did not employ such Poisson mixtures for individual level differential expression testing.

Based on the observations $\{X_{ij}^{(k)}, i \in [N_{jk}], j \in [n_k], k \in [K]\}$ as well as the measure/expression models (1.1)–(1.2), a natural question to ask is whether there exists any population-level gene expression difference among the K groups. For this, we propose to leverage a Fisher-Pitman-type permutation test based on consistent estimators $\{\widetilde{Q}_j^{(k)}, j \in [n_k], k \in [K]\}$ of the mixing distributions $\{Q_j^{(k)}, j \in [n_k], k \in [K]\}$ under Wasserstein metrics, which have received much attention in recent mixture distribution estimation literature (see, among many others, Nguyen (2013), Tian, Kong, and Valiant (2017), Vinayak et al. (2019), Wu and Yang (2020), and the references therein). Particularly appealing choices to us include the NPMLE $\widehat{Q}_j^{(k)}$ and its Poisson-smoothed one $h_{\widehat{Q}_j^{(k)}}$ (notation to be introduced by the end of this section); see Section 2 ahead for the detailed description of the testing procedure.

Many methods have been developed for differential expression analysis of scRNA-seq data (Chen, Ning, and Shi 2019). However, their focus is differential expression between two groups of cells instead of two groups of individuals. For individual level testing, a standard approach is to add up gene expression across all the cells (of a particular cell type) of an individual to create a pseudo-bulk sample, and then apply the methods for differential expression analysis using bulk RNA-seq data, such as DESeq2 (Love, Huber, and Anders 2014). The novelty of our proposed procedure is that we assess differential expression across individuals using cell level data instead of pseudo-bulk data. Furthermore, the proposed tests are shown to be consistent against their ANOVA-type alternatives, that is, they are able to asymptotically distinguish the null from any fixed alternative where the "between-group" variation is larger than the "within-group" variation, a result that sheds insight to the power of the developed tests and is in line with classic observations (Hoeffding 1952; Robinson 1973).[1]

[1] In addition to developing a more flexible nonparametric model, another route to boost the power of differential expression analysis is to de-noise the scRNA-seq data; see Zhang et al. (2022) for a proposal along that track.

As a byproduct of our theoretical study, this article further justifies the use of NPMLEs via establishing their rate-optimality in estimating the Poisson mixing distribution under the Wasserstein-1 (W_1) metric. Although the consistency of the NPMLEs has been established in the literature for different nonparametric mixture models (see Simar (1976) for nonparametric Poisson mixtures; and Chen (2017) and the references therein for more general models), NPMLEs' rates of convergence and their matching to a minimax lower bound are long standing until very recently. Built on the breakthroughs in binomial (Tian, Kong, and Valiant 2017; Vinayak et al. 2019) and Gaussian mixtures (Wu and Yang 2020) (see also Jiang and Zhang (2019) for a related study on the nonparametric likelihood ratio test) as well as the new analytical techniques devised in Jiao et al. (2015), Wu and Yang (2016), Jiao, Han, and Weissman (2018), and Han and Shiragur (2021), we are now able to further the optimality of NPMLEs to the nonparametric Poisson mixtures under minimal assumptions on the true mixing distribution function. These results yield additional theoretical support for the use of NPMLEs in our developed tests.

The rest of this article is organized as follows. Section 2 describes the model setup and studies the size and power of the proposed permutation tests. Section 3 discusses implementation of the developed test. The finite-sample performance of the developed (smoothed or not) NPMLE-based permutation tests is investigated in Section 4. Section 5 applies the studied tests to a real scRNA-seq data containing single brain nuclei from autism subjects and healthy controls (Velmeshev et al. 2019) and discover significantly differentially expressed genes that cannot be detected using the benchmark DESeq2 method applied on pseudo-bulk data (Love, Huber, and Anders 2014). In Section 6, we justify the use of NPMLEs in the permutation tests outlined in Section 2 by providing minimax optimality results for the NPMLE for nonparametric mixture of Poissons. All the proofs are relegated to a supplementary materials.

Notation. For any two distributions P, Q on the real line, the Wasserstein-1 distance is defined to be $W_1(P, Q) := \sup_{\ell \in \mathrm{Lip}_1} \int \ell(\mathrm{d}P - \mathrm{d}Q)$, where Lip_1 represents all 1-Lipschitz functions. For any distribution Q on the nonnegative real line, we define its Poisson smoothed version as

$$h_Q(x) := \int_0^\infty e^{-\lambda} \frac{\lambda^x}{x!} \mathrm{d}Q(\lambda), \; x = 0, 1, 2, \ldots$$

For any two constants a, b, we denote $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$.

2. Permutation Tests

2.1. Setup

Throughout this section, it is assumed that the observations are heterogeneous count data $\{X_{ij}^{(k)}, i \in [N_{jk}], j \in [n_k], k \in [K]\}$ with $N_{jk} = N_{jk,n} \to \infty$ and $n_k = n_{k,n} \to \infty$ as $n := \sum n_k \to \infty$. In contrast, $K \geq 2$ is assumed to be a fixed integer. It is further assumed that the probability measures $Q_j^{(k)}$'s in (1.1) have a common support $[0, B]$ for some $B > 0$ that is known a priori (see appendix Section C for a real implementation) and kept to be fixed in this section; later in Section 6 we will explore

a more general setting where $B = B_n$ is allowed to increase with n.

To facilitate the approach to distinguishing differences among the K groups, in addition to the measurement model (1.1) and the expression model (1.2), a third-layer "population model" is introduced to encourage iid randomness among each n_k within-group expression models:

$$\text{for each } k \in [K]: \quad Q_1^{(k)}, \ldots, Q_{n_k}^{(k)} \overset{\text{iid}}{\sim} \mathcal{Q}_k. \qquad \text{(population model)} \tag{2.1}$$

Here $\mathcal{Q}_k$ is understood to be a probability measure over the Prohorov-metric topology of the space of probability measures that are defined on the Borel σ-field of $[0, B]$; details about constructing Prohorov-metric topology are referred to pp. 72–73 in Billingsley (1999). Following the discussions in Sarkar and Stephens (2021, sec. "Modeling scRNA-seq data"), we do not specify $\mathcal{Q}_k$ except for assuming boundedness and well-definedness.

To wrap up, the model considered in this manuscript, summarizing the three layers ((1.1), (1.2), (2.1)), is

$$\left\{X_{ij}^{(k)}, i \in [N_{jk}], j \in [n_k], k \in [K]\right\} \text{ are independently distributed with PMFs } \int \left[\int_0^B e^{-\lambda r_{ij}^{(k)}} \frac{\{\lambda r_{ij}^{(k)}\}^x}{x!} \mathrm{d}Q(\lambda)\right] \mathrm{d}\mathcal{Q}_k(Q), \quad x = 0, 1, 2, \ldots. \tag{2.2}$$

Under the above model, it is understood that $\mathcal{Q}_1, \ldots, \mathcal{Q}_K$ and $K \geq 2$ are fixed, all of which won't change with n. Besides $\mathcal{Q}_1, \ldots, \mathcal{Q}_K$ and accordingly the random measures $Q_j^{(k)}$'s, the observations $X_{ij}^{(k)}$'s also depend on the read depths $r_{ij}^{(k)} = r_{ij,n}^{(k)}$'s that are allowed to change with n. We are hence faced with a triangular array of possibly highly heterogeneous observations.

2.2. Tests

Under Model (2.2), we are interested in testing the following null hypothesis,

$$H_0 : \mathcal{Q}_1 = \mathcal{Q}_2 = \cdots = \mathcal{Q}_K, \tag{2.3}$$

and aim to detect any population-level difference between groups. Note that here, due to the incorporation of read depths $r_{ij}^{(k)}$'s, the measurements themselves even within each group are generally not identically distributed; thus, a naive empirical distribution function based test could be substantially biased.

The main interest of this article is to explore how robust a Fisher-Pitman-type test can be when each unobserved subject-level random measure $Q_j^{(k)}$ is replaced by a plug-in-type estimate $\widetilde{Q}_j^{(k)}$ and its Poisson-smoothed version $h_{\widetilde{Q}_j^{(k)}}$ calculated from the measurements $X_{1j}^{(k)}, \ldots, X_{N_{jk}j}^{(k)}$. To this end, let's regulate $\widetilde{Q}_j^{(k)}$ as follows.

Definition 2.1. For any $j \in [n_k]$ and any $k \in [K]$, an estimator $\widetilde{Q}_j^{(k)}$ of $Q_j^{(k)}$ is said to be subject-specific conditionally W_1-consistent (shorthanded as "conditionally W_1-consistent") if it is (i) a function of $X_{1j}^{(k)}, \ldots, X_{N_{jk}j}^{(k)}$; (ii) of support $[0, B]$; and (iii) satisfying

$$E\left\{W_1\left(\widetilde{Q}_j^{(k)}, Q_j^{(k)}\right) \,\middle|\, Q_j^{(k)}\right\} \to 0 \text{ as } N_{jk} = N_{jk,n} \to \infty \tag{2.4}$$

for almost all $Q_j^{(k)}$ with regard to the measure $\mathcal{Q}_k$.

We next consider the Poisson-smoothed mixing distribution estimator

$$h_{\widetilde{Q}_j^{(k)}} := \int_0^\infty e^{-\lambda} \frac{\lambda^x}{x!} \mathrm{d}\widetilde{Q}_j^{(k)}(\lambda)$$

based on any conditionally W_1-consistent estimator $\widetilde{Q}_j^{(k)}$. It justifies the use of smoothed NPMLEs as an alternative to $\widetilde{Q}_j^{(k)}$. Note that in the classic setting when read depths are all forced to be equal, Proposition 3.1 in Lambert and Tierney (1984) showed that some $h_{\widetilde{Q}_j^{(k)}}$ can approximate $h_{Q_j^{(k)}}$ polynomially fast. Accordingly, although the differences between $Q_j^{(k)}$'s can be larger than those between $h_{Q_j^{(k)}}$'s, the differences between $\widetilde{Q}_j^{(k)}$'s can be smaller than those between $h_{\widetilde{Q}_j^{(k)}}$'s. See also Section 4.1 for some numerical results as well as Han, Miao, and Shen (2021) for some related theoretical discussions.

Theorem 2.1. Suppose $\widetilde{Q}_j^{(k)}$ is conditionally W_1-consistent. Then

$$E\left\{W_1\left(h_{\widetilde{Q}_j^{(k)}}, h_{Q_j^{(k)}}\right) \,\middle|\, Q_j^{(k)}\right\} \to 0 \text{ as } N_{jk} = N_{jk,n} \to \infty$$

for almost all $Q_j^{(k)}$ with regard to the measure $\mathcal{Q}_k$.

A particularly appealing candidate estimator of the mixing distribution is the following NPMLE $\widehat{Q}_j^{(k)}$ with read depth incorporated:

$$\widehat{Q}_j^{(k)} \in \underset{Q \text{ of support } [0,B]}{\operatorname{argmax}} \sum_{i \in [N_{jk}]} \log \int_0^\infty e^{-\lambda r_{ij}^{(k)}} \frac{\{\lambda r_{ij}^{(k)}\}^{X_{ij}^{(k)}}}{X_{ij}^{(k)}!} \mathrm{d}Q(\lambda). \tag{2.5}$$

Note that here $\widehat{Q}_j^{(k)}$ may not be unique due to read depths, and if there are multiple choices, pick any one of them (see Remark 3.1). We shall discuss the calculation of $\widehat{Q}_j^{(k)}$ in Section 3. The next theorem shows that NPMLEs are conditionally W_1-consistent under no further assumptions on the population measures $\mathcal{Q}_k$'s except for the already imposed bounded support one.

Theorem 2.2 (Conditionally W_1-consistency of NPMLEs). Assume $N_{jk} = N_{jk,n} \to \infty$ as $n \to \infty$, $r_{ij}^{(k)} = r_{ij,n}^{(k)} \in [\gamma_0, \gamma_1]$ are uniformly upper and lower bounded by two positive universal constants γ_0, γ_1, and $\mathcal{Q}_k$'s have a common fixed support $[0, B]$. We then have the NPMLEs $\widehat{Q}_j^{(k)}$'s are all conditionally W_1-consistent.

Remark 2.1. In the literature, consistency of NPMLEs of mixing distributions under the classical iid mixture distribution setup (corresponding to the case with all read depths identical to each other) has been studied in depth. Notable results include Kiefer and Wolfowitz (1956), Simar (1976), and Pfanzagl (1988); note also the survey by Chen (2017). However, although arising naturally from single-cell genomics modeling, read-depth-incorporated nonparametric mixture distributions have not received much attention in mathematical statistics and, to our knowledge, Theorem 2.2 delivers the first consistency result for NPMLEs under this heterogeneous setting.

Based on any conditionally W_1-consistent estimators $\{\widetilde{Q}_j^{(k)}\}$ of $\{Q_j^{(k)}\}$ and their Poisson-smoothed versions $h_{\widetilde{Q}_j^{(k)}}$'s, the proposed ANOVA-type (pseudo-F) test statistics are

$$\widetilde{F} := \frac{\frac{1}{n}\sum_{k_1,k_2\in[K]}\sum_{j_1\in[n_{k_1}],j_2\in[n_{k_2}]} W_1\left(\widetilde{Q}_{j_1}^{(k_1)},\widetilde{Q}_{j_2}^{(k_2)}\right)^2 - \sum_{k\in[K]}\frac{1}{n_k}\sum_{j_1,j_2\in[n_k]} W_1\left(\widetilde{Q}_{j_1}^{(k)},\widetilde{Q}_{j_2}^{(k)}\right)^2}{\sum_{k\in[K]}\frac{1}{n_k}\sum_{j_1,j_2\in[n_k]} W_1\left(\widetilde{Q}_{j_1}^{(k)},\widetilde{Q}_{j_2}^{(k)}\right)^2}$$

and

$$\widetilde{F}_h := \frac{\frac{1}{n}\sum_{k_1,k_2\in[K]}\sum_{j_1\in[n_{k_1}],j_2\in[n_{k_2}]} W_1\left(h_{\widetilde{Q}_{j_1}^{(k_1)}},h_{\widetilde{Q}_{j_2}^{(k_2)}}\right)^2 - \sum_{k\in[K]}\frac{1}{n_k}\sum_{j_1,j_2\in[n_k]} W_1\left(h_{\widetilde{Q}_{j_1}^{(k)}},h_{\widetilde{Q}_{j_2}^{(k)}}\right)^2}{\sum_{k\in[K]}\frac{1}{n_k}\sum_{j_1,j_2\in[n_k]} W_1\left(h_{\widetilde{Q}_{j_1}^{(k)}},h_{\widetilde{Q}_{j_2}^{(k)}}\right)^2}.$$

It is ready to check that these two test statistics both reduce to the original one-way ANOVA statistic if the examined space is the real space equipped with the Euclidean norm. The studied statistics then generalize the one-way ANOVA statistics to the W_1-metric measure space with different inputs (mixing distribution smoothed or not); similar generalizations have been made in various other (non-)Euclidean spaces (Anderson 2001; Mielke and Berry 2007; Petersen and Müller 2019).

We then move on to introduce the corresponding permuted ANOVA-type test statistics. To this end, for each permutation $\pi : [n] \to [n]$, let $\Pi^{j,k} = (\Pi_1^{j,k}, \Pi_2^{j,k}) := \pi^{\uparrow}(j,k)$ represent the original subject and population indices corresponding to "the jth subject in the kth group" after permutation π. The permuted test statistics are

$$\widetilde{F}^{\pi} := \frac{\frac{1}{n}\sum_{k_1,k_2\in[K]}\sum_{j_1\in[n_{k_1}],j_2\in[n_{k_2}]} W_1\left(\widetilde{Q}_{j_1}^{(k_1)},\widetilde{Q}_{j_2}^{(k_2)}\right)^2 - \sum_{k\in[K]}\frac{1}{n_k}\sum_{j_1,j_2\in[n_k]} W_1\left(\widetilde{Q}^{(\Pi_2^{j_1,k})}_{\Pi_1^{j_1,k}},\widetilde{Q}^{(\Pi_2^{j_2,k})}_{\Pi_1^{j_2,k}}\right)^2}{\sum_{k\in[K]}\frac{1}{n_k}\sum_{j_1,j_2\in[n_k]} W_1\left(\widetilde{Q}^{(\Pi_2^{j_1,k})}_{\Pi_1^{j_1,k}},\widetilde{Q}^{(\Pi_2^{j_2,k})}_{\Pi_1^{j_2,k}}\right)^2}$$

and

$$\widetilde{F}_h^{\pi} := \frac{\frac{1}{n}\sum_{k_1,k_2\in[K]}\sum_{j_1\in[n_{k_1}],j_2\in[n_{k_2}]} W_1\left(h_{\widetilde{Q}_{j_1}^{(k_1)}},h_{\widetilde{Q}_{j_2}^{(k_2)}}\right)^2 - \sum_{k\in[K]}\frac{1}{n_k}\sum_{j_1,j_2\in[n_k]} W_1\left(h_{\widetilde{Q}^{(\Pi_2^{j_1,k})}_{\Pi_1^{j_1,k}}},h_{\widetilde{Q}^{(\Pi_2^{j_2,k})}_{\Pi_1^{j_2,k}}}\right)^2}{\sum_{k\in[K]}\frac{1}{n_k}\sum_{j_1,j_2\in[n_k]} W_1\left(h_{\widetilde{Q}^{(\Pi_2^{j_1,k})}_{\Pi_1^{j_1,k}}},h_{\widetilde{Q}^{(\Pi_2^{j_2,k})}_{\Pi_1^{j_2,k}}}\right)^2}.$$

The following are the Fisher-Pitman-type permutation tests with nominal level α:

$$\widetilde{T}_{\alpha} := \begin{cases} 1, & \text{if } P(\widetilde{F}^{\pi} < \widetilde{F} \mid \widetilde{Q}_j^{(k)}\text{'s}) \geq 1-\alpha, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\widetilde{T}_{h,\alpha} := \begin{cases} 1, & \text{if } P(\widetilde{F}_h^{\pi} < \widetilde{F}_h \mid \widetilde{Q}_j^{(k)}\text{'s}) \geq 1-\alpha, \\ 0, & \text{otherwise,} \end{cases}$$

where the probability here is only with respect to the random permutation π.

As the (Poisson smoothed-)NPMLEs are chosen, the corresponding tests $\widetilde{T}_{\alpha}$ and $\widetilde{T}_{h,\alpha}$ are specified as $\widehat{T}_{\alpha}$ and $\widehat{T}_{h,\alpha}$.

2.3. Theory

This section provides the necessary theoretical support on the presented tests $\widetilde{F}^{\pi}$ and $\widetilde{F}_h^{\pi}$. Particular focus is on the asymptotic size and consistency against Robinson-type ANOVA alternatives (see Theorem 3 in Robinson 1973). To minimize assumptions and for presentation clearness, we are focused on the following balanced design case:

Assumption 2.1. The design is balanced so that $n_k = n/K$ and $N_{jk} = N$ for $j \in [n_k]$, $k \in [K]$. In addition, it is assumed that the sets $\{r_{ij}^{(k)}, i \in [N]\}$ are invariant with respect to both j and k.

Remark 2.2. We note that Assumption 2.1 can be weakened in a straightforward manner to allow for $n_k/n \to 1/K$, N_{jk}'s asymptotically comparable, and the sets $\{r_{ij}^{(k)}, i \in [N_{jk}]\}$ all weakly converge to a same probability measure that does not depend on the particular choice of j and k (see Shi, Drton, and Han (2022, Proposition 2.2) as well as Deb and Sen (in press) for a similar setup in the recent independence testing literature). We however do not pursue these tracks but rather leave them to the readers of interest to verify.

Our first result concerns with the sizes of proposed tests, is still valid with a finite sample size, and is a direct consequence of a long line of literature on permutation-based tests.

Theorem 2.3 (Size validity). We have, for any finite N and n, as long as H_0 in (2.3) and Assumption 2.1 hold,

$$P(\widetilde{T}_\alpha = 1|H_0) \le \alpha \quad \text{and} \quad P(\widetilde{T}_{h,\alpha} = 1|H_0) \le \alpha.$$

In the following, we are focused on asymptotic results with the balanced design and let $N = N_n \to \infty$ as $n \to \infty$. The next theorem is the main result of this section.

Theorem 2.4 (Power against ANOVA-type fixed alternatives). Consider $\widetilde{Q}_j^{(k)}$'s to be conditionally W_1-consistent estimators of $Q_j^{(k)}$'s. If Assumption 2.1 holds, then the following two statements are true.

(a) Under any fixed alternative regarding $\mathcal{Q}_1, \ldots, \mathcal{Q}_K$ such that

$$H_1 : \frac{1}{K}\sum_{k\in[K]} E\left\{W_1\left(Q_1^{(k)}, Q_2^{(k)}\right)^2\right\} < \sum_{k_1\neq k_2\in[K]} \frac{E\{W_1(Q_1^{(k_1)}, Q_1^{(k_2)})^2\}}{K(K-1)}, \tag{2.6}$$

we have $\lim_{n\to\infty} P(\widetilde{T}_\alpha = 1|H_1) = 1$ for each $\alpha \in (0,1)$.

(b) Under any fixed alternative regarding $\mathcal{Q}_1, \ldots, \mathcal{Q}_K$ such that

$$H_{1,h} : \frac{1}{K}\sum_{k\in[K]} E\left\{W_1\left(h_{Q_1^{(k)}}, h_{Q_2^{(k)}}\right)^2\right\} < \sum_{k_1\neq k_2\in[K]} \frac{E\left\{W_1\left(h_{Q_1^{(k_1)}}, h_{Q_1^{(k_2)}}\right)^2\right\}}{K(K-1)}, \tag{2.7}$$

we have $\lim_{n\to\infty} P(\widetilde{T}_{h,\alpha} = 1|H_{1,h}) = 1$ for each $\alpha \in (0,1)$.

Remark 2.3. Assumption (2.6) states that the average of the Wasserstein distances within groups is less than the average of the Wasserstein distances between groups. If then, our theory suggested that one is able to test the differences between groups using the permuted ANOVA-type test statistics. Assumption (2.7) is analogous, while since the test statistics is based on the smoothed NPMLEs, assumptions have to be made on $(h_{Q_1^{(k)}}, h_{Q_2^{(k)}})$'s.

Specific to (smoothed-)NPMLEs, the following theorem is a direct consequence of Theorems 2.2–2.4.

Corollary 2.1. Suppose Assumption 2.1 and all conditions in Theorem 2.2 hold. Then the following are true for any $\alpha \in (0,1)$.

(a) For any finite N and n, as long as H_0 in (2.3) holds, we have

$$P(\widehat{T}_\alpha = 1|H_0) \le \alpha \quad \text{and} \quad P(\widehat{T}_{h,\alpha} = 1|H_0) \le \alpha.$$

(b) Concerning any fixed alternative H_1 (or H_{1h}), we have

$$\lim_{n\to\infty} P(\widehat{T}_\alpha = 1 \mid H_1) = 1 \quad \text{and}$$
$$\lim_{n\to\infty} P(\widehat{T}_{h,\alpha} = 1 \mid H_{1,h}) = 1.$$

Remark 2.4. We note that, in both Theorem 2.4 and Corollary 2.1, the considered alternatives are not the complement of H_0 in (2.3); in particular, there exist $\mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_K$ such that (a) the probability measures are different, yet (b) they yield the same within- and between-group distances. For such alternatives, it is obvious that the tests considered in Theorem 2.4 and Corollary 2.1 have no power, and hence they are not consistent tests of H_0. This lack of consistency is well known in the ANOVA literature (see chap. 7.3 in Lehmann and Romano 2005).

3. Algorithms

This section presents three algorithms to calculate (2.5),

1. the vertex direction method (VDM), see Fedorov (1972), Simar (1976), Wu (1978a, 1978b), Böhning (1982), and Lindsay (1983a);
2. the vertex exchange method (VEM), see Böhning (1985, 1986);
3. the intra simplex direction method (ISDM), see Lesperance and Kalbfleisch (1992).

Some more recent algorithmic developments along this track can be found in, for example, Groeneboom, Jongbloed, and Wellner (2008) and Koenker and Mizera (2014).

To simplify the notation, in this section we remove j, k from the subscript and use $\{X_i, i \in [N]\}$ and $\{r_i, i \in [N]\}$ to denote the sample points and the corresponding read-depths. Moreover, we use $\widehat{Q}$ to denote the NPMLE defined in (2.5) based on $\{X_i, i \in [N]\}$ and $\{r_i, i \in [N]\}$. For a discrete measure G on $[0, B]$ with support points $\{\lambda_m, m \in [M]\}$, let $G(\lambda_m)$ stand for the mass G assigned at λ_m for each $m \in [M]$. We define

$$\Phi(G) := \frac{1}{N}\sum_{i\in[N]} \log\left(\sum_{m\in[M]} G(\lambda_m) e^{-\lambda_m r_i} (\lambda_m r_i)^{X_i}\right)$$

and its directional derivative from G to δ_λ as

$$\Phi'(G, \delta_\lambda) := \lim_{\epsilon\to 0^+} \epsilon^{-1}\left\{\Phi\{(1-\epsilon)G \oplus \epsilon\delta_\lambda\} - \Phi(G)\right\} = \frac{1}{N}\sum_{i\in[N]} \frac{e^{-\lambda r_i}(\lambda r_i)^{X_i}}{\sum_{m\in[M]} G(\lambda_m) e^{-\lambda_m r_i}(\lambda_m r_i)^{X_i}} - 1.$$

Here δ_λ represents the unit measure at $\lambda \in [0, B]$. Lastly, for any two signed measures ν_1 and ν_2 on the real line, we denote $\nu_1 \oplus \nu_2$ as the sum of ν_1 and ν_2, and $\nu_1 \ominus \nu_2$ as the sum of ν_1 and $-\nu_2$.

With these notation, we are now ready to present the VDM, VEM, and ISDM algorithms for calculating $\widehat{Q}$.

The VDM Algorithm

Step 0 (*Initialization*). Select a point $\lambda_1 \in (0, B]$. Let $G_1 = \delta_{\lambda_1}$ be the initial value. Set the loop index $L = 1$.

Step 1. If $\max_{\lambda\in[0,B]} \Phi'(G_L, \delta_\lambda) = 0$, then stop and return G_L. Otherwise, find $\lambda_{\max} = \operatorname{argmax}_{\lambda\in[0,B]} \Phi'(G_L, \delta_\lambda)$.

Step 2. Find $\alpha_{\max} = \operatorname{argmax}_{\alpha\in[0,1]} \Phi\Big\{(1-\alpha)G_L \oplus \alpha\delta_{\lambda_{\max}}\Big\}$.

Step 3. Set $G_{L+1} = (1-\alpha)G_L \oplus \alpha_{\max}\delta_{\lambda_{\max}}$. Set $L = L+1$ and go to Step 1.

The VEM Algorithm

Step 0 (*Initialization*). Select a point $\lambda_1 \in (0, B]$. Let $G_1 = \delta_{\lambda_1}$ be the initial value. Set the loop index $L = 1$.

Step 1. If $\max_{\lambda\in[0,B]} \Phi'(G_L, \delta_\lambda) = 0$, then stop and return G_L. Otherwise, find $\lambda_{\max} = \operatorname{argmax}_{\lambda\in[0,B]} \Phi'(G_L, \delta_\lambda)$ and $\lambda_{\min} = \operatorname{argmin}_{\lambda\in\operatorname{supp}(G_L)} \Phi'(G_L, \delta_\lambda)$, where $\operatorname{supp}(G_L)$ stands for the support of G_L.

Step 2. Find $\alpha_{\max} = \operatorname{argmax}_{\alpha\in[0,1]} \Phi\Big\{G_L \oplus \Big(\alpha G_L(\lambda_{\min})(\delta_{\lambda_{\max}} \ominus \delta_{\lambda_{\min}})\Big)\Big\}$.

Step 3. Set $G_{L+1} = G_L \oplus \Big(\alpha_{\max} G_L(\lambda_{\min})(\delta_{\lambda_{\max}} \ominus \delta_{\lambda_{\min}})\Big)$. Set $L = L+1$ and go to Step 1.

The ISDM Algorithm

Step 0 (*Initialization*). Select a point $\lambda_1 \in (0, B]$. Let $G_1 = \delta_{\lambda_1}$ be the initial value. Set the loop index $L = 1$.

Step 1. If $\max_{\lambda\in[0,B]} \Phi'(G_L, \delta_\lambda) = 0$, then stop and return G_L. Otherwise, find all local maxima $\lambda_{\max,1}, \ldots, \lambda_{\max,\mathcal{N}}$ of $\lambda \mapsto \Phi'(G_L, \delta_\lambda)$ on $[0, B]$, where $\mathcal{N}$ represents the number of local maxima.

Step 2. Find $(\alpha_{\max,0}, \ldots, \alpha_{\max,\mathcal{N}}) = \operatorname{argmax}_{\alpha_0,\ldots,\alpha_{\mathcal{N}}} \Phi\Big\{(1-\alpha_0)G_L \oplus \alpha_1\delta_{\lambda_{\max,1}} \oplus \cdots \oplus \alpha_{\mathcal{N}}\delta_{\lambda_{\max,\mathcal{N}}}\Big\}$ subject to $\alpha_0 \geq 0, \alpha_1 \geq 0, \ldots, \alpha_{\mathcal{N}} \geq 0$ and $\alpha_0 + \alpha_1 + \cdots + \alpha_{\mathcal{N}} = 1$.

Step 3. Set $G_{L+1} = (1-\alpha_{\max,0})G_L \oplus \alpha_{\max,1}\delta_{\lambda_{\max,1}} \oplus \cdots \oplus \alpha_{\max,\mathcal{N}}\delta_{\lambda_{\max,\mathcal{N}}}$. Set $L = L+1$ and go to Step 1.

The convergence of VDM, VEM, and ISDM is guaranteed by the following theorem.

Theorem 3.1. Assuming $r_i > 0$ for each $i \in [N]$. For each of VDM, VEM, and ISDM, if it stops for some L, then we have $\Phi(G_L) = \Phi(\widehat{Q})$; otherwise, $\Phi(G_L) \to \Phi(\widehat{Q})$ as $L \to \infty$.

Remark 3.1. Unlike in the traditional setting where all read depths are identical, when heterogeneous read depths are incorporated, although $G \mapsto \Phi(G)$ is still a concave function, there is no theoretical guarantee about the uniqueness of $\widehat{Q}$'s that maximize the objective function and whether the maximizer is unique or not is still open. This issue of computational uniqueness shall be compared to the parallel result in Theorem 2.2, which provides theoretical guarantee for the consistency of an arbitrary maximizer of the objective function as the sample size increases to infinity.

4. Simulation Studies

This section is split to two parts. The first part aims to show that the two NPMLE-based tests presented in Section 2 cannot dominate each other. To this end, we fix $K = 2$ and consider three designs with several cases of population models that will be detailed in Section 4.1. The second part provides some preliminary discussions on the computation complexity of the proposed algorithm.

4.1. Finite-Sample Experiments

Designs.

(A) Balanced designs with all read depths set to be 1, $n_1 = n_2 = 10$, and $N_{jk} = 50$, 100, and 500 for each j, k.

(B) Balanced designs with read-depth effects with $n_1 = n_2 = 10$ and $N_{jk} = 50$, 100 and 500 for each j, k. In addition, in each round of the simulation, $\{r_{i1}^{(1)}, i \in [N_{11}]\}$ are iid generated from Uniform(0.5, 1.5) and then let $r_{ij}^{(k)} = r_{i1}^{(1)}$ for each j, k.

(C) A particular unbalanced design motivated by the single-cell RNA-seq data in Section 5 ahead, with $n_1 = 10, n_2 = 13$ and N_{jk} be as in Table 1. For each round of the simulation, $\{r_{ij}^{(k)}, i \in [N_{jk}], j \in [n_k], k \in [K]\}$ are iid generated from Uniform(0.5, 1.5).

We then move on to specify the population model (2.1) used in our simulation studies. Hereafter, let $\text{Gam}(a, b; B)$ denote a truncated Gamma distribution with a shape parameter $a > 0$, a rate parameter $b > 0$, and with any realization larger than B shrunken to B. Let $\{\Delta_j^{(k)}, j \in [n_k], k \in [K]\}$ be iid generated from Uniform$(-1, 1)$.

Population models.

1. (a) $Q_j^{(k)} \sim \text{Gam}(14+\Delta_j^{(k)}, 7/4; 50)$ for each $j \in [n_k], k \in [2]$.
   (b) $Q_j^{(k)} \sim \text{Gam}(14+\Delta_j^{(k)}, 7; 50)$ for each $j \in [n_k], k \in [2]$.
   (c) $Q_j^{(k)} \sim \text{Gam}(6+\Delta_j^{(k)}, 1; 50)$ for each $j \in [n_k], k \in [2]$.
2. (a) $Q_j^{(1)} \sim \text{Gam}(14+\Delta_j^{(1)}, 7/4; 50)$ for $j \in [n_1]$ and $Q_j^{(2)} \sim \text{Gam}(6+\Delta_j^{(2)}, 3/4; 50)$ for $j \in [n_2]$.
   (b) $Q_j^{(1)} \sim \text{Gam}(14+\Delta_j^{(1)}, 7/3; 50)$ for $j \in [n_1]$ and $Q_j^{(2)} \sim \text{Gam}(6+\Delta_j^{(2)}, 1; 50)$ for $j \in [n_2]$.
   (c) $Q_j^{(1)} \sim \text{Gam}(14+\Delta_j^{(1)}, 7/2; 50)$ for $j \in [n_1]$ and $Q_j^{(2)} \sim \text{Gam}(6+\Delta_j^{(2)}, 3/2; 50)$ for $j \in [n_2]$.

Table 1. N_{jk} in the unbalanced design (Design (C)).

$N_{1,1}$	$N_{2,1}$	$N_{3,1}$	$N_{4,1}$	$N_{5,1}$	$N_{6,1}$	$N_{7,1}$	$N_{8,1}$	$N_{9,1}$	$N_{10,1}$	$N_{1,2}$	$N_{2,2}$
388	1142	162	391	215	278	284	193	542	106	202	759
$N_{3,2}$	$N_{4,2}$	$N_{5,2}$	$N_{6,2}$	$N_{7,2}$	$N_{8,2}$	$N_{9,2}$	$N_{10,2}$	$N_{11,2}$	$N_{12,2}$	$N_{13,2}$	
415	69	327	431	414	451	275	733	422	65	362	

3. (a) $Q_j^{(1)} \sim \text{Gam}(4+\Delta_j^{(1)},1;20)$ for $j \in [n_1]$ and $Q_j^{(2)} \sim \text{Gam}(5+\Delta_j^{(2)},1;20)$ for $j \in [n_2]$.
   (b) $Q_j^{(1)} \sim \text{Gam}(5+\Delta_j^{(1)},1;20)$ for $j \in [n_1]$ and $Q_j^{(2)} \sim \text{Gam}(6+\Delta_j^{(2)},1;20)$ for $j \in [n_2]$.
   (c) $Q_j^{(1)} \sim \text{Gam}(6+\Delta_j^{(1)},1;20)$ for $j \in [n_1]$ and $Q_j^{(2)} \sim \text{Gam}(7+\Delta_j^{(2)},1;20)$ for $j \in [n_2]$.
4. (a) $Q_j^{(1)} \sim \text{Gam}(11+\Delta_j^{(1)},1;50)$ for $j \in [n_1]$ and $Q_j^{(2)} \sim \text{Gam}(12+\Delta_j^{(2)},1;50)$ for $j \in [n_2]$.
   (b) $Q_j^{(1)} \sim \text{Gam}(12+\Delta_j^{(1)},1;50)$ for $j \in [n_1]$ and $Q_j^{(2)} \sim \text{Gam}(13+\Delta_j^{(2)},1;50)$ for $j \in [n_2]$.
   (c) $Q_j^{(1)} \sim \text{Gam}(13+\Delta_j^{(1)},1;50)$ for $j \in [n_1]$ and $Q_j^{(2)} \sim \text{Gam}(14+\Delta_j^{(2)},1;50)$ for $j \in [n_2]$.

Our focus is on examining as well as comparing the empirical performance of the tests $\widehat{T}_\alpha$ and $\widehat{T}_{h,\alpha}$ with NPMLE calculated using the oracle B. Both of them are based on an exact critical value approximated by 1000 Monte Carlo simulations. The underlying nominal significance level is 0.05. For each setting, 1000 rounds of simulations were performed. We use VEM to compute NPMLEs with a stop tolerance 0.01. Optimization in Step 1 and Step 2 in VEM is implemented by the default interior-point algorithm in Matlab; see the support page of function "fmincon" for further details.

Table 2 shows the empirical sizes and powers (rejection frequencies) of tests $\widehat{T}_\alpha$ and $\widehat{T}_{h,\alpha}$. In short, the results confirm our earlier theoretical claims on the sizes and powers of T_α and $\widehat{T}_{h,\alpha}$ in the different models and balanced designs (Designs (A) and (B)). Moreover, even under the unbalanced design (Design C), T_α and $\widehat{T}_{h,\alpha}$ still perform well in terms of their empirical sizes and powers.

Some more detailed comparisons between T_α and $\widehat{T}_{h,\alpha}$ are in line. The following observations depend on the "signal strengths" D and D_h, defined as follows:

$$D := E\{W_1(Q_1^{(1)}, Q_1^{(2)})^2\} - \left(E\{W_1(Q_1^{(1)}, Q_2^{(1)})^2\} + E\{W_1(Q_1^{(2)}, Q_2^{(2)})^2\}\right)/2 \tag{4.1}$$

and

$$D_h := E\{W_1(h_{Q_1^{(1)}}, h_{Q_1^{(2)}})^2\} - \left(E\{W_1(h_{Q_1^{(1)}}, h_{Q_2^{(1)}})^2\} + E\{W_1(h_{Q_1^{(2)}}, h_{Q_2^{(2)}})^2\}\right)/2. \tag{4.2}$$

First, empirical results for Model 1 illustrates that under H_0, empirical powers are close to the nominal level $\alpha = 0.05$, confirming the size validity of $\widehat{T}_\alpha$ and $\widehat{T}_{h,\alpha}$. In addition, even under the unbalanced design (Design C), empirical powers are stable and close to the nominal level $\alpha = 0.05$, indicating the robustness of the studied tests.

Second, we compare the empirical powers using Models 2, 3, and 4. In Model 2, D is significantly larger than D_h and the corresponding empirical powers of $\widehat{T}_\alpha$ are all larger than these of $\widehat{T}_{h,\alpha}$ in all three considered designs (Designs (A), (B), and (C)). This phenomenon is not surprising to us as the difference between variation between groups and variation within groups in mixing distributions is much larger than that in mixture distributions. Therefore, $\widehat{T}_\alpha$ is more powerful than $\widehat{T}_{h,\alpha}$.

In Model 3, D is approximately equal to D_h and the empirical power of $\widehat{T}_\alpha$ is smaller than the empirical power of $\widehat{T}_{h,\alpha}$ when N is small (e.g., 50 and 100). However, the empirical powers of $\widehat{T}_\alpha$ and $\widehat{T}_{h,\alpha}$ are close when N is large. Similar observation applies to Model 4, where D is also approximately equal to D_h. However, compared to Model 3, the mixing distributions in Model 4 have larger B and thus the empirical powers of $\widehat{T}_{h,\alpha}$ are higher than the empirical powers of $\widehat{T}_\alpha$ even for $N = 500$, especially under

Table 2. Empirical sizes and powers of $\widehat{T}_\alpha$ and $\widehat{T}_{h,\alpha}$; here D and D_h are defined in (4.1) and (4.2).

Model	1(a)	1(b)	1(c)	2(a)	2(b)	2(c)	3(a)	3(b)	3(c)	4(a)	4(b)	4(c)
D	0	0	0	0.59	0.32	0.15	0.99	0.99	0.99	0.99	0.99	0.99
D_h	0	0	0	0.22	0.10	0.03	0.99	0.99	0.99	0.99	0.99	0.99
N	Empirical sizes/powers for $\widehat{T}_\alpha$ under Design (A)											
50	0.054	0.050	0.045	0.644	0.595	0.356	0.811	0.772	0.698	0.538	0.501	0.502
100	0.043	0.055	0.053	0.901	0.835	0.583	0.870	0.872	0.843	0.723	0.680	0.668
500	0.049	0.049	0.060	0.996	0.999	0.965	0.952	0.958	0.941	0.850	0.831	0.829
N	Empirical sizes/powers for $\widehat{T}_{h,\alpha}$ under Design (A)											
50	0.054	0.045	0.049	0.284	0.210	0.111	0.833	0.816	0.767	0.650	0.635	0.624
100	0.038	0.063	0.049	0.371	0.264	0.138	0.896	0.892	0.882	0.797	0.788	0.771
500	0.042	0.047	0.055	0.492	0.309	0.186	0.951	0.961	0.947	0.944	0.924	0.921
N	Empirical sizes/powers for $\widehat{T}_\alpha$ under Design (B)											
50	0.044	0.048	0.058	0.644	0.508	0.338	0.796	0.763	0.729	0.559	0.522	0.520
100	0.053	0.050	0.062	0.863	0.779	0.518	0.878	0.862	0.846	0.714	0.735	0.679
500	0.036	0.052	0.054	1.000	0.998	0.972	0.958	0.952	0.939	0.922	0.920	0.913
N	Empirical sizes/powers for $\widehat{T}_{h,\alpha}$ under Design (B)											
50	0.044	0.050	0.054	0.262	0.193	0.100	0.821	0.806	0.772	0.632	0.619	0.602
100	0.058	0.041	0.053	0.350	0.276	0.132	0.885	0.877	0.858	0.772	0.788	0.759
500	0.036	0.045	0.057	0.501	0.414	0.187	0.956	0.950	0.943	0.932	0.928	0.924
N	Empirical sizes/powers for $\widehat{T}_\alpha$ under Design (C)											
Table 1	0.048	0.050	0.051	0.994	0.988	0.900	0.962	0.940	0.951	0.910	0.904	0.907
N	Empirical sizes/powers for $\widehat{T}_{h,\alpha}$ under Design (C)											
Table 1	0.047	0.051	0.052	0.452	0.346	0.173	0.966	0.947	0.952	0.929	0.920	0.922

Design (A). Some pilot studies to explain this phenomenon will be put in Section 6, where we analyze the finite-sample behavior of the NPMLE under an exploratory simplified setting where all read depths are fixed to be 1. There, the rate of convergence of NPMLE, at the worst case, is showed to be $O(\log\log N/\log N)$; in contrast, Lambert and Tierney (1984, Lemma 4.1 and Theorem 4.1) showed that the Poisson-smoothed NPMLE attains a near-root-n rate of convergence to the mixture distribution.

4.2. Time Complexity and Actual Running Time

This section provides some discussions on the algorithm implemented in Section 4.1, with B assumed to be bounded. This algorithm consists of two major parts: (a) implementing the VEM algorithm detailed in Section 3 for calculating estimates of the mixing distributions; (b) feeding the estimates to the permutation tests in Section 2.2. In the following we discussed the computation complexity of these two parts separately.

4.2.1. Time Complexity of NPMLE

We start with an analysis of the VEM algorithm. Using the notation of Section 3, every iteration of VEM involves the following three steps:

1. find $\lambda_{\max} = \text{argmax}_{\lambda\in[0,B]}\,\Phi'(G_L,\delta_\lambda)$;
2. find $\lambda_{\min} = \text{argmin}_{\lambda\in\text{supp}(G_L)}\,\Phi'(G_L,\delta_\lambda)$;
3. find $\alpha_{\max} = \text{argmax}_{\alpha\in[0,1]}\,\Phi\{G_L \oplus [\alpha G_L(\lambda_{\min})(\delta_{\lambda_{\max}} \ominus \delta_{\lambda_{\min}})]\}$,

where it is reminded that

$$\Phi'(G_L,\delta_\lambda) = \frac{1}{N}\sum_{i\in[N]}\frac{e^{-\lambda r_i}(\lambda r_i)^{X_i}}{\sum_{m\in[M]} G_L(\lambda_m)\, e^{-\lambda_m r_i}(\lambda_m r_i)^{X_i}} - 1.$$

In Step 1, to obtain an ϵ-accuracy (with respect to the objective function) solution, we search over an ϵ-grid on $[0,B]$ (i.e., $[0,\epsilon,2\epsilon,\ldots,B]$) and then use a gradient-descent algorithm with the best result over the grid as the initial value to obtain an accurate solution, an idea commonly used in literatures; see, for example, Lindsay (1995, chap. 6.1). Since the derivative of $\lambda\mapsto\Phi'(G_L,\delta_\lambda)$ is bounded on $[0,B]$, it is immediate that this grid search method indeed leads to an ϵ-accuracy solution for sufficiently small ϵ; here the boundedness of the derivative of $\lambda\mapsto\Phi'(G_L,\delta_\lambda)$ follows from the derivative of $\lambda\mapsto e^{-\lambda r_i}(\lambda r_i)^{X_i}$, that is,

$$\frac{e^{-r_i\lambda}(r_i\lambda)^{X_i}(X_i - r_i\lambda)}{\lambda} = \begin{cases} -r_i e^{-r_i\lambda} & \text{if } X_i = 0,\\ r_i^{X_i} e^{-r_i\lambda}\lambda^{X_i-1}(X_i - r_i\lambda) & \text{if } X_i \geq 1,\end{cases}$$

which is bounded on $[0,B]$. The time complexity for the grid search method is $O(N^2/\epsilon)$, where N^2 comes from the sum of index $i\in[N]$ and $m\in[M]$ with $M\leq N$. For the time complexity of the gradient-descent algorithm, it follows from Nesterov's Theorem (Nesterov 2003, Theorem 2.1.14) that one needs $O(N^2/\epsilon)$ iterations to obtain an ϵ-accuracy solution as long as the objective function has a Lipschitz continuous gradient, or equivalently, the boundedness of the second derivative of $\lambda\mapsto\Phi'(G_L,\delta_\lambda)$. This follows from the boundedness of the second derivative of $\lambda\mapsto e^{-\lambda r_i}(\lambda r_i)^{X_i}$, that is,

$$\frac{e^{-\lambda r_i}(\lambda r_i)^{X_i}\left(\lambda^2 r_i^2 - X_i(2\lambda r_i + 1) + X_i^2\right)}{\lambda^2} = \begin{cases} r_i^2 e^{-\lambda r_i} & \text{if } X_i = 0,\\ r_i^2 e^{-\lambda r_i}(\lambda r_i - 2) & \text{if } X_i = 1,\\ r_i^{X_i} e^{-\lambda r_i}\lambda^{X_i-2}[(\lambda r_i - X_i)^2 - X_i] & \text{if } X_i \geq 2.\end{cases}$$

on $[0,B]$.

In Step 2, since the support size of G_L is at most N and the summation over i is from 1 to N in $\Phi'(G_L,\delta_\lambda)$, a brutal force method requires at most $O(N^2)$ to find the $\lambda_{\min}$. The time complexity for Step 3 is the same as for Step 1 by analogous arguments, and hence the total time complexity in three steps is $O(N^2/\epsilon)$ to obtain an ϵ-accuracy solution in each iteration.

To determine the total time complexity, it remains to determine how many iterations (recalling that each iteration contains the above three steps) are needed. It follows from Böhning (1982, Assumption (iii) and its proof) or Equation (A.11) in the supplementary materials that the increase of the objective function is strict and linear after each iteration. Accordingly, the number of iterations is $O(1/\epsilon)$ to obtain an ϵ-accuracy solution and the total time complexity for the VEM algorithm is $O(N^2/\epsilon^2)$.

Table 3 shows the actual running time (in seconds) for computing NPMLEs averaged over 100 simulations; recall that the tolerance level is set to be 0.01. Here we adopt the Design (B) with $N = 500$, $\{r_i : i\in[N]\} \overset{\text{iid}}{\sim}\text{Uniform}(0.5,1.5)$, and consider three population models, Gam(14,1;50), Gam(10,1;50), and Gam(6,1;50). The simulation is conducted over a laptop with a 1.8 GHz Intel Core i5 processor and an 8 GB memory.

Table 3. Actual running time (in seconds) for computing NPMLEs.

The mixing distribution Q	Gam(14,1;50)	Gam(10,1;50)	Gam(6,1;50)
Actual running time	6.31	5.74	4.99

4.2.2. Time Complexity of Permutation Tests

We then move on to examine the time complexity of the remaining parts. For computing W_1 distance between two Poisson mixture distributions, where the corresponding mixing distributions are both supported over at most M points, first note that the time complexity of computing the mixture density at any point is $O(M)$. Moreover, it follows from Poisson tail inequality (see Lemma B.5 in the supplementary materials) that as long as the support mixing distributions is bounded by some positive constant, it suffices to compute the mixture densities over at most $O(\sqrt{\log(1/\epsilon)})$ points to obtain an ϵ-accuracy solution. As a consequence, the time complexity for computing the W_1 distance between two Poisson mixtures is $O(M\sqrt{\log(1/\epsilon)})$.

With a little abuse of notation, let's use N to denote $\max_{i,k}\{N_{jk}\}$. Since each NPMLE is supported over at most N points, the time complexity of computing the W_1 distance of estimated mixture and mixing distributions is $O(N\sqrt{\log(1/\epsilon)})$ and hence the time complexity of computing the W_1 distance matrix is $O(n_1 n_2 N\sqrt{\log(1/\epsilon)})$, where n_1 and n_2 are the number of subjects in each group. As a result, the total time complexity

of performing permutation-based test is

$$O(n_1 n_2 N\sqrt{\log(1/\epsilon)} + n_1 n_2 \mathcal{M}),$$

where $\mathcal{M}$ represents the set number of permutations.

For an example of the actual running time, with $\mathcal{M} = 1000$, the total time to perform permutation tests with mixture and mixing distribution estimates input is averagely 0.19 sec on a laptop with a 1.8 GHz Intel Core i5 processor and a 8 GB memory, where the mixing distributions are NPMLEs estimated under Design (A) and Model 4(c) with $N = 500$.

5. Applications to Single-Cell Genomics

This section applies the studied permutation tests to a scRNA-seq data. There has been a large literature studying fitting RNA-seq data using Poisson mixtures including, for example, over-dispersed Poisson model (Robinson, McCarthy, and Smyth 2010), Poisson-Gamma model (Love, Huber, and Anders 2014; Huang et al. 2018), Poisson-Beta model (Vu et al. 2016), Poisson-log normal model (Silva et al. 2019), Poisson mixture model with K-clusters (Rau et al. 2015), finite Poisson mixture models (Wu, Qin, and Zhu 2013), zero-inflated mixture Poisson linear models (Liu, Jiang, and Yu 2019), Poisson mixture models with unimodal mixing distributions (Lu 2018). Compared to parametric Poisson mixture models, nonparametric Poisson mixture models haven't received much attention; some notable exceptions include Bi and Davuluri (2013), Dadaneh, Qian, and Zhou (2018), Sarkar and Stephens (2021), the latter of which was closely followed by us.

5.1. Dataset Description

The scRNA-seq data used in this article is obtained from Velmeshev et al. (2019), which focused on autism spectrum disorder (ASD) and recorded gene expression of 23 subjects (13 ASD vs. 10 control) and 18,041 genes for each subject from 17 different cell types and 2 different brain regions. Here we focus on the brain region prefrontal cortex, which is more relevant to autism disease etiology. Moreover, each subject has seven covariates including age, sex, diagnosis, capbatch, seqbatch, postmortem interval (PMI), and RNA integrity number (RIN).

We focus on a pre-selected subset including 100 genes (names of the genes put in Table 4) that were documented to be related to body height; for relation between ASD and body height, see, for example, Fukumoto et al. (2011) and Chawarska et al. (2011). In addition to permutation testing with either estimated mixing distributions or mixture distributions, we also consider DESeq2 (Love, Huber, and Anders 2014) as a benchmark. In implementing the two considered permutation tests, we adopt a common strategy to incorporate four covariates: age, sex, seqbatch, and RIN. The other two covariates PMI and capbatch are not significantly associated with gene expression given the other covariates, since their p-value distributions across all genes are uniform. The corresponding tests were denoted as $\widehat{T}_Z$ (with the original NPMLE) and $\widehat{T}_{h,Z}$ (with the Poisson-smoothed NPMLE). Implementation details—including the choice of B, the choice of read depths, and an additional step of covariate adjustment based on the work of Zhang et al. (2022)—were put in the Section C, supplementary materials.

5.2. Implementation Results

Using $\widehat{T}_Z$, 9 genes are significant under the threshold of false discovery rate (FDR) 0.05 after multiple testing correction by the Benjamini-Hochberg procedure. Replacing $\widehat{T}_Z$ by $\widehat{T}_{h,Z}$, 8 genes are significant under the same threshold of FDR and 7 genes are coincident with significant genes found by $\widehat{T}_Z$. This shows some consistency between $\widehat{T}_Z$ and $\widehat{T}_{h,Z}$.

Furthermore, by DESeq2 there are seven significant genes under the same threshold of FDR and all of them are coincident with significant genes found by $\widehat{T}_Z$. In other words, among significant genes found by $\widehat{T}_Z$, 78% significant genes are coincident with genes found by DESeq2 and 22% are new which means $\widehat{T}_Z$ could enrich the set of significant genes found by the standard method DESeq2.

Similarly, six genes are coincident with significant genes found by $\widehat{T}_{h,Z}$. In other words, among significant genes found by $\widehat{T}_{h,Z}$, 75% significant genes are coincident with genes found by DESeq2 and 25% are new which means $\widehat{T}_{h,Z}$ could enrich the set of significant genes found by the standard method DESeq2. In one word, both $\widehat{T}_Z$ and $\widehat{T}_{h,Z}$ could enrich the set of significant genes found by DESeq2. Further details are summarized in Figure 1.

Our results can also be justified by functions of significant genes. For example, fasting blood glucose measurement is not only one of functions of gene DHRS7B, but also related to ASD (Hoirisch-Clapauch and Nardi 2019). More such results are summarized in Table 5.

Table 4. All genes used in Section 5.

DST	CHSY3	TSC2	EHD4	HERC1	KIF16B	DLGAP1	PIK3CG
ELL	ODF2L	FBXL5	LNX1	ERGIC3	CBFA2T2	FAM20A	STAT2
DAP	SSH2	WDR60	SAXO1	FOXP2	SAMD4A	TSPAN9	ARAP3
GHR	KCNK9	RGL1	SOCS5	ZNF76	ADAMTS2	DHRS7B	PNMA8C
KIZ	SHPRH	RBMS3	MFSD2B	NR4A3	CCDC171	RAB33A	WDR70
IL16	MTMR3	CDK10	ZNF628	CAPZB	ATXN7L3	PSKH1	FGFRL1
BST2	UMAD1	CPED1	ESYT2	LRRC43	SMARCA4	MYO18A	IL17RD
LHX2	FBP2	ZC3H13	SRRM2	NOTCH1	HSD17B3	SBNO1	EIF3H
RLF	LAYN	SUSD5	DOT1L	WARS2	RPS4XP13	PHF11	CDK11B
DAZL	CYFIP2	ST7L	CWC27	C9orf152	TOB1-AS1	HIF1AN	KLHL28
BCL9	LRWD1	LMO7	PTENP1	CEP112	LINC01572	PPP4R2	UBE2Z
NRK	GCLC	PPM1H	ITGA9	HIP1R	PPP1R16A	POLR3E	TANC2
ANKDD1A		ZNF710-AS1		ZRANB2-AS2		DNAJC27-AS1	

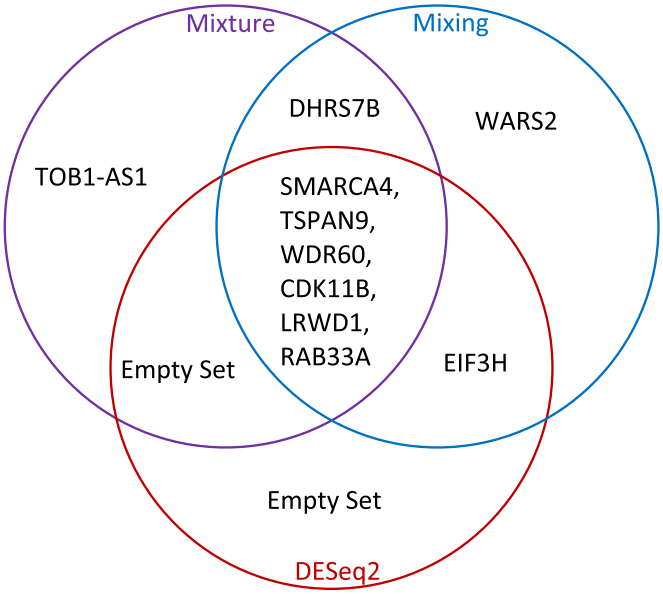

Figure 1. Significant genes selected using Mixing ($\widehat{T}_Z$), Mixture ($\widehat{T}_{h,Z}$), and DESeq2 methods.

Table 5. Significant genes on ASD with some literature support.

Gene name	Related functions	Literatures
DHRS7B	Fasting blood glucose	Hoirisch-Clapauch and Nardi (2019)
WDR60	Abnormality of refraction	Ezegwui et al. (2014)
EIF3H	Reaction time measurement	Baisch et al. (2017)
LRWD1	Insomnia measurement	Hohn et al. (2019)
RAB33A	Bipolar disorder	Joshi et al. (2012)
TSPAN9	Creatinine measurement	Cameron et al. (2017)
WARS2, CDK11B	Heel bone mineral density	Calarge and Schlechte (2017)
SMARC4, TOB1-AS1	Cholesterol measurement	Benachenhou et al. (2019)

NOTE: The first column includes names of genes, the second column includes functions potentially related to ASD, and the third column includes literature support.

6. Minimax Optimality of the Poisson NPMLEs

This section provides additional theoretical support for the use of NPMLEs in forming up the tests $\widehat{T}_\alpha$ and $\widehat{T}_{h,\alpha}$ in Section 2. To this end, due to the technical challenges, focus is restricted to a simplified setting of (2.2), where the observations $\{X_i, i \in [N]\}$ independently follow a distribution of PMF

$$h_Q(x) = \int_0^B e^{-\lambda}\frac{\lambda^x}{x!}\mathrm{d}Q(\lambda), \quad x = 0, 1, 2, \ldots, \tag{6.1}$$

where Q is a deterministic measure supported on $[0, B]$ that cannot be characterized by a simple parametric model. This is exactly the classic nonparametric Poisson mixture setup, and we study the nonasymptotic behavior of the following NPMLE

$$\widehat{Q} = \underset{Q \text{ of support } [0,B]}{\operatorname{argmax}} \sum_{i\in[N]} \log h_Q(X_i). \tag{6.2}$$

Note that the above NPMLE is the simplified version of (2.5) with all read depths there forced to be one.

There has been an enormous literature studying the NPMLE (6.2) under the nonparametric Poisson mixture model (6.1). Earlier results on the existence, discreteness (of the NPMLE support), and computation include, among many others, Simar (1976), Laird (1978), Jewell (1982), Lindsay (1983a), Lindsay (1983b), and Lindsay and Roeder (1993); see also Lindsay (1995) for a survey. Consistency of NPMLEs were established in, among many others, Kiefer and Wolfowitz (1956), Simar (1976), and Pfanzagl (1988); see also Chen (2017) for a survey.

Beyond these important results, there has been another track of substantial research that is focused on establishing the minimax rate in estimating the mixing distribution (mostly on the density function) of nonparametric Poisson mixtures. Notable results there include Zhang (1995), Loh and Zhang (1996), van de Geer (1996), Hengartner (1997), van de Geer (2003), Roueff and Rydén (2005), and Rebafka and Roueff (2015). However, to our knowledge, a study on the minimax optimality and the corresponding convergence rates for NPMLEs under a fully nonparametric Poisson mixture model is still absent from the literature.

Before presenting our main result in this section, we would love to highlight again that, due to the nature of nonasymptotic analysis, all the parameters in the model (6.1), including B, are allowed to change with N. This is a strict generalization of the "asymptotic" setting in Section 2, where, due to the additional hardness of handling the read depth as well as for simplifying notation and assumptions, we do not intend to establish similar nonasymptotic results.

Our first theorem concerns with the NPMLE's rate of convergence.

Theorem 6.1 (Upper bound of NPMLEs).

(a) Suppose there exists a universal constant $c_0 > 0$ such that $B \le c_0 \log N$. Then there exists a positive constant $C = C(c_0)$ such that for all sufficiently large N ($> N_0(c_0)$) we have

$$\sup_{Q \text{ of support } [0,B]} E\Big\{W_1(\widehat{Q}, Q)\Big\} \le C\frac{B}{\log N}\log\left(\frac{\log N}{B} \vee e\right). \tag{6.3}$$

(b) Suppose there exist universal strictly positive constants c_0, C_0 and $\epsilon_0 \in (0, 1/3)$ such that $B \in [c_0 \log N, C_0 N^{1/3-\epsilon_0}]$. Then there exists a strictly positive constant $C = C(\epsilon_0, c_0)$ such that for all sufficiently large N ($> N_0(c_0, C_0, \epsilon_0)$) we have

$$\sup_{Q \text{ of support } [0,B]} E\Big\{W_1(\widehat{Q}, Q)\Big\} \le C\sqrt{\frac{B}{\log N}}. \tag{6.4}$$

Remark 6.1. We believe that the condition $B \le C_0 N^{1/3-\epsilon_0}$ in Theorem 6.1(b) is somewhat necessary. In particular, supposing there exist $c_0 > 0, \delta > 0$ such that $B \ge c_0 N^{2+\delta}$, we conjecture that a sufficiently small constant $c = c(c_0, \delta)$ exists such that for all N large enough, $\inf_{\widetilde{Q}} \sup_Q E\{W_1(\widetilde{Q}, Q)\} \ge c\sqrt{B}$ but not $\sqrt{B/\log N}$; here the infimum and supremum are taken over all estimators and all distributions of support $[0, B]$. At this moment, we do not know how to prove this conjecture.

Our second theorem concerns with minimax lower bounds in estimating mixing distributions in model (6.1). Combined with Theorem 6.1, it confirms the NPMLE's minimax optimality.

Theorem 6.2 (Minimax lower bound of mixing distribution estimation).

(a) Supposing there exists $c_0 > 0$ such that $B \le c_0 \log N$, it follows that for any $N \ge 3$,

$$\inf_{\widetilde{Q}} \sup_{Q} E\{W_1(\widetilde{Q}, Q)\} \ge \frac{B}{24e \log N} \log\left(\frac{16 c_0 \log N}{B}\right).$$

(b) Supposing there exists $c_0 > 0$ such that $B \ge c_0 \log N$, it follows that for any $N \ge 1$,

$$\inf_{\widetilde{Q}} \sup_{Q} E\{W_1(\widetilde{Q}, Q)\} \ge \frac{3}{40e^4} \sqrt{\frac{B}{c_0 \log N}}.$$

In the above, the infimum and supremum are understood to be taken over all estimators and all distributions of support $[0, B]$.

Remark 6.2. Under fully nonparametric binomial mixture models, minimax optimal convergence rates for NPMLEs of mixing distributions were obtained by Vinayak et al. (2019, sec. 3) in terms of the W_1 distance. Under fully nonparametric binomial and Gaussian mixture models, Tian, Kong, and Valiant (2017, Theorem 1) and Wu and Yang (2020, p. 1985) obtained optimal convergence rates for moment-based estimators in terms of W_1 distance; see also Polyanskiy and Wu (2020, Remark 2). Nguyen (2013, Theorems 1 and 2) upper bounded the Wasserstein distance between mixing distributions by the divergence between the corresponding mixture distributions under general mixture models, with normal mixture models as an example in Example 2. However, their results cannot be applied here since Theorem 1 restricts the mixing distribution being discrete and Theorem 2 is only for convolution mixture models.

Remark 6.3. Comparing the theoretical results in Theorem 6.1 to the empirical observations in Sections 4 and 5 suggests an interesting discrepancy, that a slow logarithmic convergence rate of estimation can yield tests of power in finite-sample studies. To explain this, we note that the derived rates of convergence are in the minimax sense and hence necessarily conservative against many special types of alternatives. In particular, recent results (Saha and Guntuboyina 2020; Kim and Guntuboyina 2022) revealed that, in various related scenarios, NPMLEs are provably able to adapt to the structure of mixing density and hence yield much faster convergence rates. A complete study of this phenomenon in our setting, however, has to be left to the future.

Supplementary Materials

The supplemental materials contain all the proofs and the implementation details for the real experiment presented in Section 5.

Acknowledgments

The authors would like to thank the associate editor and two referees for their very helpful comments and suggestions. In particular, the authors would like to thank the referees for pointing out a mistake in the original submission as well as providing the authors with an elegant approach to simplifying the proofs. They would also like to thank Yihong Wu for his very informative remarks on the issue of uniqueness of NPMLEs and concavity of the nonparametric Poisson likelihood functions, and Jiahua Chen, Matthew Stephens, and Jon Wellner for pointing out related literature and for helpful discussions.

References

Anderson, M. J. (2001), "A New Method for Non-parametric Multivariate Analysis of Variance," *Austral Ecology*, 26, 32–46. [394,397]

Baisch, B., Cai, S., Li, Z., and Pinheiro, V. (2017), "Reaction Time of Children with and without Autistic Spectrum Disorders," *Open Journal of Medical Psychology*, 6, 166–178. [403]

Benachenhou, S., Etcheverry, A., Galarneau, L., Dubé, J., and Çaku, A. (2019), "Implication of Hypocholesterolemia in Autism Spectrum Disorder and its Associated Comorbidities: A Retrospective Case–Control Study," *Autism Research*, 12, 1860–1869. [403]

Berry, K. J., and Mielke, P. W. (1983), "Moment Approximations as an Alternative to the F Test in Analysis of Variance," *British Journal of Mathematical and Statistical Psychology*, 36, 202–206. [394]

Bi, Y., and Davuluri, R. V. (2013), "NPEBseq: Nonparametric Empirical Bayesian-based Procedure for Differential Expression Analysis of RNA-seq Data," *BMC Bioinformatics*, 14, 262. [402]

Billingsley, P. (1999), *Convergence of Probability Measures* (2nd ed.), New York: Wiley. [396]

Böhning, D. (1982), "Convergence of Simar's Algorithm for Finding the Maximum Likelihood Estimate of a Compound Poisson Process," *Annals of Statistics*, 10, 1006–1008. [398,401]

——— (1985), "Numerical Estimation of a Probability Measure," *Journal of Statistical Planning and Inference*, 11, 57–69. [398]

——— (1986), "A Vertex-Exchange-Method in D-Optimal Design Theory," *Metrika*, 33, 337–347. [398]

Boik, R. J. (1987), "The Fisher-Pitman Permutation Test: A Non-Robust Alternative to the Normal Theory F Test when Variances are Heterogeneous," *British Journal of Mathematical and Statistical Psychology*, 40, 26–42. [394]

Calarge, C. A., and Schlechte, J. A. (2017), "Bone Mass in Boys with Autism Spectrum Disorder," *Journal of Autism and Developmental Disorders*, 47, 1749–1755. [403]

Cameron, J. M., Levandovskiy, V., Roberts, W., Anagnostou, E., Scherer, S., Loh, A., and Schulze, A. (2017), "Variability of Creatine Metabolism Genes in Children with Autism Spectrum Disorder," *International Journal of Molecular Sciences*, 18, 1665. [403]

Chawarska, K., Campbell, D., Chen, L., Shic, F., Klin, A., and Chang, J. (2011), "Early Generalized Overgrowth in Boys with Autism," *Archives of General Psychiatry*, 68, 1021–1031. [402]

Chen, G., Ning, B., and Shi, T. (2019), "Single-Cell RNA-Seq Technologies and Related Computational Data Analysis," *Frontiers in Genetics*, 10, 317. [395]

Chen, J. (2017), "Consistency of the MLE under Mixture Models," *Statistical Science*, 32, 47–63. [395,397,403]

Chung, E., and Romano, J. P. (2013), "Exact and Asymptotically Robust Permutation Tests," *Annals of Statistics*, 41, 484–507. [394]

Dadaneh, S. Z., Qian, X., and Zhou, M. (2018), "BNP-seq: Bayesian Nonparametric Differential Expression Analysis of Sequencing Count Data," *Journal of the American Statistical Association*, 113, 81–94. [402]

Deb, N., and Sen, B. (in press), "Multivariate Rank-Based Distribution-Free Nonparametric Testing Using Measure Transportation," *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2021.1923508. [398]

Ezegwui, I., Lawrence, L., Aghaji, A., Obiekwe, O., Okoye, O., Onwasigwe, E., and Ebigbo, P. (2014), "Refractive Errors in Children with Autism in a Developing Country," *Nigerian Journal of Clinical Practice*, 17, 467–470. [403]

Fedorov, V. (1972), *Theory of Optimal Experiments Designs*, New York: Academic Press. [398]

Fisher, R. A. (1925), *Statistical Methods for Research Workers*, Edinburgh: Oliver and Boyd. [394]

——— (1935), *Design of Experiments*, Edinburgh: Oliver and Boyd. [394]

Fukumoto, A., Hashimoto, T., Mori, K., Tsuda, Y., Arisawa, K., and Kagami, S. (2011), "Head Circumference and Body Growth in Autism Spectrum Disorders," *Brain and Development*, 33, 569–75. [402]

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2008), "The Support Reduction Algorithm for Computing Non-parametric Function Estimates in Mixture Models," *Scandinavian Journal of Statistics*, 35, 385–399. [398]

Han, F., Miao, Z., and Shen, Y. (2021), "Nonparametric Mixture MLEs under Gaussian-Smoothed Optimal Transport Distance," arXiv preprint arXiv:2112.02421. [396]

Han, Y., and Shiragur, K. (2021), "On the Competitive Analysis and High Accuracy Optimality of Profile Maximum Likelihood," in *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1317–1336. [395]

Hengartner, N. W. (1997), "Adaptive Demixing in Poisson Mixture Models," *Annals of Statistics*, 25, 917–928. [403]

Hoeffding, W. (1952), "The Large-Sample Power of Tests based on Permutations of Observations," *Annals of Mathematical Statistics*, 23, 169–192. [394,395]

Hohn, V. D., de Veld, D., Mataw, K., van Someren, E., and Begeer, S. (2019), "Insomnia Severity in Adults with Autism Spectrum Disorder is Associated with Sensory Hyper-Reactivity and Social Skill Impairment," *Journal of Autism and Developmental Disorders*, 49, 2146–2155. [403]

Hoirisch-Clapauch, S., and Nardi, A. (2019), "Autism Spectrum Disorders: Let's Talk about Glucose?" *Translational Psychiatry*, 9, 51. [402,403]

Huang, M., Wang, J., Torre, E., Dueck, H., Shaffer, S., Bonasio, R., Murray, J., Raj, A., Li, M., and Zhang, N. R. (2018), "SAVER: Gene Expression Recovery for UMI-based Single Cell RNA Sequencing," *Nature Methods*, 15, 539–542. [402]

Jewell, N. P. (1982), "Mixtures of Exponential Distributions," *Annals of Statistics*, 10, 479–484. [403]

Jiang, W., and Zhang, C.-H. (2019), "Rate of Divergence of the Nonparametric Likelihood Ratio Test for Gaussian Mixtures," *Bernoulli*, 25, 3400–3420. [395]

Jiao, J., Han, Y., and Weissman, T. (2018), "Minimax Estimation of the L_1 Distance," *IEEE Transactions on Information Theory*, 64, 6672–6706. [395]

Jiao, J., Venkat, K., Han, Y., and Weissman, T. (2015), "Minimax Estimation of Functionals of Discrete Distributions," *IEEE Transactions on Information Theory*, 61, 2835–2885. [395]

Joshi, G., Biederman, J., Petty, C., Goldin, R. L., Furtak, S. L., and Wozniak, J. (2012), "Examining the Comorbidity of Bipolar Disorder and Autism Spectrum Disorders: A Large Controlled Analysis of Phenotypic and Familial Correlates in a Referred Population of Youth with Bipolar I Disorder with and without Autism Spectrum Disorders," *Journal of Clinical Psychiatry*, 74, 578–586. [403]

Kiefer, J., and Wolfowitz, J. (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *Annals of Mathematical Statistics*, 27, 887–906. [397,403]

Kim, A. K., and Guntuboyina, A. (2022), "Minimax Bounds for Estimating Multivariate Gaussian Location Mixtures," *Electronic Journal of Statistics*, 16, 1461–1484. [404]

Koenker, R., and Mizera, I. (2014), "Convex Optimization, Shape Constraints, Compound Decisions, and Empirical Bayes Rules," *Journal of the American Statistical Association*, 109, 674–685. [398]

Laird, N. (1978), "Nonparametric Maximum Likelihood Estimation of a Mixing Distribution," *Journal of the American Statistical Association*, 73, 805–811. [403]

Lambert, D., and Tierney, L. (1984), "Asymptotic Properties of Maximum Likelihood Estimates in the Mixed Poisson Model," *Annals of Statistics*, 12, 1388–1399. [396,401]

Lehmann, E. L., and Romano, J. P. (2005), *Testing Statistical Hypotheses* (Vol. 3), Springer. [398]

Lesperance, M. L., and Kalbfleisch, J. D. (1992), "An Algorithm for Computing the Nonparametric MLE of a Mixing Distribution," *Journal of the American Statistical Association*, 87, 120–126. [398]

Lindsay, B. G. (1983a), "The Geometry of Mixture Likelihoods: A General Theory," *Annals of Statistics*, 11, 86–94. [398,403]

——— (1983b), "The Geometry of Mixture Likelihoods, Part II: The Exponential Family," *Annals of Statistics*, 11, 783–792. [403]

——— (1995), "Mixture Models: Theory, Geometry and Applications," *NSF-CBMS Regional Conference Series in Probability and Statistics*, 5, I–163. [401,403]

Lindsay, B. G., and Roeder, K. (1993), "Uniqueness of Estimation and Identifiability in Mixture Models," *Canadian Journal of Statistics*, 21, 139–147. [403]

Liu, S., Jiang, Y., and Yu, T. (2019), "Modelling RNA-Seq Data with a Zero-Inflated Mixture Poisson Linear Model," *Genetic Epidemiology*, 43, 786–799. [402]

Loh, W.-L., and Zhang, C.-H. (1996), "Global Properties of Kernel Estimators for Mixing Densities in Discrete Exponential Family Models," *Statistica Sinica*, 6, 561–578. [403]

Love, M., Huber, W., and Anders, S. (2014), "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2," *Genome Biology*, 15, 550. [395,402]

Lu, M. (2018), *Generalized Adaptive Shrinkage Methods and Applications in Genomics Studies*, Chicago, IL: University of Chicago. [402]

Marascuilo, L. A., and McSweeney, M. (1977), *Nonparametric and Distribution-Free Methods for the Social Sciences*, Monterey, CA: Brooks/Cole Publishing Company. [394]

Mielke, P. W., and Berry, K. J. (2007), *Permutation Methods: A Distance Function Approach*, New York: Springer. [394,397]

Mielke Jr., P. (1984), "34 Meteorological Applications of Permutation Techniques based on Distance Functions," in *Handbook of Statistics* (Vol. 4), pp. 813–830, Amsterdam: Elsevier. [394]

Mielke Jr., P. W., Berry, K. J., and Johnson, E. S. (1976), "Multi-Response Permutation Procedures for a priori Classifications," *Communications in Statistics-Theory and Methods*, 5, 1409–1424. [394]

Nesterov, Y. (2003), *Introductory Lectures on Convex Optimization: A Basic Course*, New York: Springer. [401]

Nguyen, X. (2013), "Convergence of Latent Mixing Measures in Finite and Infinite Mixture Models," *Annals of Statistics*, 41, 370–400. [395,404]

Petersen, A., and Müller, H.-G. (2019), "Fréchet Regression for Random Objects with Euclidean Predictors," *Annals of Statistics*, 47, 691–719. [394,397]

Pfanzagl, J. (1988), "Consistency of Maximum Likelihood Estimators for Certain Nonparametric Families, in Particular: Mixtures," *Journal of Statistical Planning and Inference*, 19, 137–158. [397,403]

Pitman, E. J. G. (1938), "Significance Tests which may be Applied to Samples from any Populations III. The Analysis of Variance Test," *Biometrika*, 29, 322–335. [394]

Polyanskiy, Y., and Wu, Y. (2020), "Self-Regularizing Property of Nonparametric Maximum Likelihood Estimator in Mixture Models," arXiv preprint arXiv:2008.08244. [404]

Rau, A., Maugis-Rabusseau, C., Martin-Magniette, M.-L., and Celeux, G. (2015), "Co-Expression Analysis of High-Throughput Transcriptome Sequencing Data with Poisson Mixture Models," *Bioinformatics*, 31, 1420–1427. [402]

Rebafka, T., and Roueff, F. (2015), "Nonparametric Estimation of the Mixing Density Using Polynomials," *Mathematical Methods of Statistics*, 24, 200–224. [403]

Robinson, J. (1973), "The Large-Sample Power of Permutation Tests for Randomization Models," *Annals of Statistics*, 1, 291–296. [394,395,397]

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010), "edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data," *Bioinformatics*, 26, 139–140. [402]

Roueff, F., and Rydén, T. (2005), "Nonparametric Estimation of Mixing Densities for Discrete Distributions," *Annals of Statistics*, 33, 2066–2108. [403]

Saha, S., and Guntuboyina, A. (2020), "On the Nonparametric Maximum Likelihood Estimator for Gaussian Location Mixture Densities with Application to Gaussian Denoising," *Annals of Statistics*, 48, 738–762. [404]

Sarkar, A. K., and Stephens, M. (2021), "Separating Measurement and Expression Models Clarifies Confusion in Single Cell RNA-Seq Analysis," *Nature Genetics*, 53, 770–777. [394,395,396,402]

Scheffé, H. (1959), *The Analysis of Variance*, New York: Wiley. [394]

Shi, H., Drton, M., and Han, F. (2022), "Distribution-Free Consistent Independence Tests via Center-Outward Ranks and Signs," *Journal of the American Statistical Association*, 117, 395–410. [398]

Silva, A., Rothstein, S. J., McNicholas, P. D., and Subedi, S. (2019), "A Multivariate Poisson-Log Normal Mixture Model for Clustering Transcriptome Sequencing Data," *BMC Bioinformatics*, 20, 394. [402]

Simar, L. (1976), "Maximum Likelihood Estimation of a Compound Poisson Process," *Annals of Statistics*, 4, 1200–1209. [395,397,398,403]

Still, A., and White, A. (1981), "The Approximate Randomization Test as an Alternative to the F Test in Analysis of Variance," *British Journal of Mathematical and Statistical Psychology*, 34, 243–252. [394]

Tian, K., Kong, W., and Valiant, G. (2017), "Learning Populations of Parameters," in *Advances in Neural Information Processing Systems* (Vol. 30). [395,404]

van de Geer, S. (1996), "Rates of Convergence for the Maximum Likelihood Estimator in Mixture Models," *Journal of Nonparametric Statistics*, 6, 293–310. [403]

van de Geer, S. (2003), "Asymptotic Theory for Maximum Likelihood in Nonparametric Mixture Models," *Computational Statistics and Data Analysis*, 41, 453–464. [403]

Velmeshev, D., Schirmer, L., Jung, D., Haeussler, M., Perez, Y., Mayer, S., Bhaduri, A., Goyal, N., Rowitch, D. H., and Kriegstein, A. R. (2019), "Single-Cell Genomics Identifies Cell Type–Specific Molecular Changes in Autism," *Science*, 364, 685–689. [395,402]

Vinayak, R. K., Kong, W., Valiant, G., and Kakade, S. (2019), "Maximum Likelihood Estimation for Learning Populations of Parameters," in *International Conference on Machine Learning* (Vol. 97), pp. 6448–6457. [395,404]

Vu, T. N., Wills, Q. F., Kalari, K. R., Niu, N., Wang, L., Rantalainen, M., and Pawitan, Y. (2016), "Beta-Poisson Model for Single-Cell RNA-Seq Data Analyses," *Bioinformatics*, 32, 2128–2135. [402]

Wu, C.-F. (1978a), "Some Algorithmic Aspects of the Theory of Optimal Designs," *Annals of Statistics*, 6, 1286–1301. [398]

——— (1978b), "Some Iterative Procedures for Generating Nonsingular Optimal Designs," *Communications in Statistics-Theory and Methods*, 7, 1399–1412. [398]

Wu, H., Qin, Z., and Zhu, Y. (2013), "PM-Seq: Using Finite Poisson Mixture Models for RNA-Seq Data Analysis and Transcript Expression Level Quantification," *Statistics in Biosciences*, 5, 71–87. [402]

Wu, Y., and Yang, P. (2016), "Minimax Rates of Entropy Estimation on Large Alphabets via Best Polynomial Approximation," *IEEE Transactions on Information Theory*, 62, 3702–3720. [395]

——— (2020), "Optimal Estimation of Gaussian Mixtures via Denoised Method of Moments," *Annals of Statistics*, 48, 1981–2007. [395,404]

Zhang, C.-H. (1995), "On Estimating Mixing Densities in Discrete Exponential Family Models," *Annals of Statistics*, 23, 929–945. [403]

Zhang, M., Liu, S., Miao, Z., Han, F., Gottardo, R., and Sun, W. (2022), "Individual Level Differential Expression Analysis for Single Cell RNA-Seq Data," *Genome Biology*, 23, 1–11. [395,402]

Zhang, M. J., Ntranos, V., and Tse, D. (2020), "Determining Sequencing Depth in a Single-Cell RNA-Seq Experiment," *Nature Communications*, 11, 774. [395]