Sparse-limit approximation for t-statistics

Micól Tresoldi[1], Daniel Xiang[2] and Peter McCullagh[3]

[1] *The Dow Chemical Company, e-mail:* micol91@hotmail.com

[2] *Department of Statistics, University of Chicago, e-mail:* dxiang@uchicago.edu

[3] *Department of Statistics, University of Chicago, e-mail:* petermccullagh52@gmail.com

Abstract: In a range of genomic applications, it is of interest to quantify the evidence that the signal at site i is active given conditionally independent replicate observations summarized by the sample mean and variance $(\bar{Y}, s^2)$ at each site. We study the version of the problem in which the signal distribution is sparse, and the error distribution has an unknown site-specific variance so that the null distribution of the standardized statistic is Student-t rather than Gaussian. The main contribution of this paper is a sparse-mixture approximation to the non-null density of the t-ratio. This formula demonstrates the effect of low degrees of freedom on the Bayes factor, or the conditional probability that the site is active. We illustrate some differences on a HIV dataset for gene-expression data.

1. Introduction

We consider the sparse signal plus replicated Gaussian noise model on the real line, with $m \geq 2$ conditionally independent replicate observations at each site. By sufficiency, the observations at one site may be reduced to the sample mean and variance in the implied signal-plus-noise model

$$\bar{Y} = X + \bar{\varepsilon}, \qquad (m-1)s^2 \sim \sigma^2 \chi^2_{m-1}. \tag{1}$$

The signal X is assumed to be independent of the error, which is distributed as $N(0, \sigma^2/m)$ with variance inversely proportional to the sample size. Since the model has an unknown site-specific variance parameter, it is natural first to reduce the information at each site to the standardized ratio $T_i = m^{1/2}\bar{Y}_i/s_i$, where s_i^2 is the sample variance. The null distribution is Student-t on $m - 1$ degrees of freedom, independent for each site, and independent of σ^2. The evidence for signal activity at site i is then based on the observed value T_i in relation to the null distribution and to the values at all other sites.

This model for $(\bar{Y}, s^2)$ typically arises where there is a need to summarize data with parallel structure for the sites, e.g. in high-throughput biology settings such as microarrays ([3], [4], [8]). A common assumption in the signal detection literature is to suppose that X is distributed according to a sparse 'atom and slab' mixture distribution [12] $X \sim \pi_0 \delta_0 + (1 - \pi_0)G_1$ with null probability

$\pi_0 \approx 1$, and non-null component G_1. In this note, we work with a more general concept of sparsity [11] and derive an approximation to the marginal density of the t-ratio when the distribution of X is sparse. To this end, up to a re-scaling of X, it suffices to consider a special case of (1) with $\sigma = 1$.

In section 2 we review the definition of statistical sparsity and derive a mixture representation for the distribution of the t-ratio when the signal is sparse. In section 3 we compute conditional probabilities using this approximation and make a comparison with the probability integral transform on a real dataset of gene expression levels from HIV patients. Section 4 concludes with a brief discussion, and appendix A contains proofs of our results.

2. Statistical sparsity and t-ratios

2.1. Sparse limit

The definition of statistical sparsity proposed by [11] aims to organize the family of signal distributions by the resulting signal-plus-standard Gaussian model $X + \varepsilon$, where $\varepsilon \sim N(0, 1)$ is independent of X. In other words, two sequences (P_ν) and (Q_μ) have the same sparse limit if and only if there is a 1–1 matching $\nu \mapsto \mu(\nu)$ such that the convolutions $P_\nu * N(0, 1)$ and $Q_{\mu(\nu)} * N(0, 1)$ are hard to distinguish as $\nu \to 0$. This criterion is an equivalence relation, which is formalized in terms of weak convergence of the signal distribution over a suitably large class of integrands, the constraints of which are related to particularities of the Gaussian convolutional model. Each equivalence class is associated with an exceedance measure as follows.

Definition 2.1 ([11])**.** *A family of distributions (P_ν) indexed by $\nu > 0$ has a sparse limit with rate ρ_ν and exceedance measure $H \neq 0$ if*

$$\lim_{\nu \to 0} \rho_\nu^{-1} \int_{\mathbb{R}} w(x) \, P_\nu(\mathrm{d}x) = \int_{\mathbb{R}} w(x) \, H(\mathrm{d}x) < \infty \tag{2}$$

for every function $w : \mathbb{R} \to \mathbb{R}$ that is bounded, continuous and $O(x^2)$ at the origin.

For every $\lambda > 0$, the pair $(\lambda \rho_\nu, \lambda^{-1} H)$ is equivalent to (ρ_ν, H) in (2). To eliminate this indeterminacy, a non-zero measure H may be replaced with a unit measure such that

$$\int_{\mathbb{R}} (1 - e^{-x^2/2}) \, H(\mathrm{d}x) = 1. \tag{3}$$

This particular normalization arises naturally in the signal-plus-Gaussian noise convolution. If the family is sparse with unit exceedance measure H, then the integral

$$\int_{\mathbb{R}} (1 - e^{-x^2/2}) \, P_\nu(\mathrm{d}x) = \rho_\nu + o(\rho) \tag{4}$$

determines the rate parameter to first order in $\rho \equiv \rho_\nu$. Here, $b(\rho) = o(\rho)$ means $\frac{b(\rho)}{\rho} \to 0$ as $\rho \to 0$.

Every distribution has its own sparsity parameter or sparsity rate as defined by ([4]). By contrast, the exceedance measure is a characteristic of certain families. Thus, two sparse families $(P_\nu), (P'_\mu)$ may be put in 1–1 correspondence by their sparsity rates ρ, ρ' such that $\rho(\nu) = \rho'(\mu)$. In the sense of w-integrals occurring in ([2]), such matched pairs (P_ν, Q_μ) are indistinguishable to first order in the sparse limit if the two families have the same exceedance measure. First-order equivalence of the convolution implies an equivalence relationship on sparse families.

Definition 2.2. *Two sparse families are first-order equivalent if they have the same exceedance measure.*

2.2. Scale family

Since this paper is concerned with the effect of unknown or estimated scale parameters, it is natural to require both the family of signal distributions and the family of error distributions to be closed under scalar multiplication.

Definition 2.3. *A family of distributions P_ν indexed by $\nu > 0$ is closed under scalar multiplication if, to each pair $\nu, \sigma > 0$ there corresponds a number $\nu' > 0$ such that $X \sim P_\nu$ implies $\sigma X \sim P_{\nu'}$. A closed family is called a scale family if the group action is transitive with a single orbit, i.e., to each ν, ν' there corresponds a number $\sigma > 0$ such that $X \sim P_\nu$ implies $\sigma X \sim P_{\nu'}$. By convention, the scale family is indexed by the scale parameter σ. The family has a sparse limit if ([2]) is satisfied in the small-scale limit $\sigma \to 0$.*

A family that is closed may contain several orbits. The Student-t scale family on d degrees of freedom has one orbit for each $d > 0$. The sparsity rate and exceedance measure for the Student-t scale family are stated below, and shown to satisfy ([2]) in appendix [A].

Example 2.1 ([11]). *Let $0 < d < 2$ be given. The Student-t family on d degrees of freedom with scale parameter $\sigma > 0$ has a sparse limit as $\sigma \to 0$ with rate*

$$\rho = \frac{d^{d/2}\Gamma\left(\frac{d+1}{2}\right)}{C_d\sqrt{\pi}\Gamma(d/2)} \cdot \sigma^d.$$

The inverse-power exceedance measure

$$H_d(\mathrm{d}x) \coloneqq C_d \cdot \frac{\mathrm{d}x}{|x|^{d+1}}, \qquad C_d \coloneqq \frac{d\,2^{d/2-1}}{\Gamma(1-d/2)}$$

satisfies ([3]). In particular, for $d = 1$, $P_\sigma = C(\sigma)$ is the Cauchy distribution with probable error σ; the sparsity rate is $\rho = \sigma\sqrt{2/\pi}$, and the normalized exceedance measure is $\mathrm{d}x/(x^2\sqrt{2\pi})$.

Example 2.2. *The zero-mean Gaussian family $P_\nu = N(0, \nu^2)$ is a scale family. However, there does not exist a measure H satisfying (2), so this family does not have a sparse limit. For the same reason, neither the scaled Student-t family on $d \geq 2$ degrees of freedom nor the Laplace scale family with density $\sigma^{-1} e^{-|x|/\sigma}/2$ has a sparse limit as $\sigma \to 0$.*

Example 2.3. *The atom-and-slab Cauchy mixture*

$$P_\sigma = 0.8\delta_0 + 0.2C(\sigma)$$

is a symmetric scale family, as is the atom-free spike-and-slab mixture

$$P_\sigma = 0.8N(0, \sigma^2) + 0.2C(\sigma).$$

In both cases, (2) is satisfied as $\nu = \sigma \to 0$ with rate $0.2\sigma\sqrt{2/\pi}$ and inverse-square exceedance density $\mathrm{d}x/(x^2\sqrt{2\pi})$. Both mixtures are first-order equivalent to the sparse Cauchy model $C(\sigma)$ and to the horseshoe model [2] with density $\sigma^{-1} \log(1 + \sigma^2 x^{-2})/(2\pi)$.

The occurrence of the Student-t family indexed by $0 < d < 2$ and $\sigma > 0$ as an instance of a sparse scale family is unrelated to the principal topic of this paper, which is the distributional effect of internal normalization by the sample standard deviation. In that case, the null distribution of $T = \bar{Y}\sqrt{m}/s$ is Student-t on $m-1$ degrees of freedom, where the sample size $m \geq 2$ is arbitrarily large and $\sigma = 1$. The non-null distribution is derived in section 2.6.

The main purpose of this section is to characterize the totality of sparse scale families by the set of equivalence classes. Surprisingly, the Student-t family in Example 2.1 covers the entire range. In other words, every sparse scale family has an inverse-power exceedance measure with some index $0 < d < 2$. It follows that each equivalence class has a Student t_d representative. A precise statement is given below and a short proof is provided in appendix A.

Theorem 2.1. *Every symmetric sparse scale family has a rate function $\rho(\sigma) = \sigma^d L(\sigma^{-1})$ and an inverse-power exceedance measure $C_d \, \mathrm{d}x/|x|^{d+1}$ for some $0 < d < 2$ and slowly varying function $L : (0, \infty) \to (0, \infty)$.*

The claim that $\rho(\sigma) = \sigma^d$ for scale families, which occurs at the end of section 2.2 in [11], is incorrect, as the following example demonstrates.

Example 2.4. *The family of unimodal symmetric distributions*

$$P_\sigma(\mathrm{d}x) = \frac{\sigma \log(1 + x^2/\sigma^2)\,\mathrm{d}x}{2\pi\, x^2}$$

is a scale family whose tails are slightly heavier than Cauchy. For $K \simeq 4.3552$, similar remarks apply to the log-weighted Cauchy family

$$Q_\sigma(\mathrm{d}x) = \frac{\sigma \log(1 + x^2/\sigma^2)\,\mathrm{d}x}{K\,(\sigma^2 + x^2)},$$

except that Q_σ has modes at $\pm\sigma\sqrt{e-1}$, with zero density at the origin. Never-theless, both families satisfy the limit condition (2) in the form

$$\lim_{\sigma\to 0}\frac{P_\sigma(\mathrm{d}x)}{-\sigma\log\sigma} = \frac{\mathrm{d}x}{\pi x^2}, \qquad \lim_{\sigma\to 0}\frac{Q_\sigma(\mathrm{d}x)}{-\sigma\log\sigma} = \frac{2\,\mathrm{d}x}{Kx^2},$$

with $L(\sigma^{-1}) \propto -\log\sigma$ in Theorem 2.1. Accordingly, both are sparse with rate $\rho \propto -\sigma\log\sigma$, and inverse-square exceedance density. Despite the difference in rate functions, they belong to the same equivalence class (Student t_1) as all families in Example 2.3.

2.3. Signal plus noise convolution

Let X be a random signal with distribution P_ν taken from a family of symmetric distributions having a sparse limit with rate ρ and unit exceedance measure H. The observation Y is generated from the signal-plus-standard Gaussian model,

$$Y = X + \varepsilon, \qquad \varepsilon \sim N(0,1).$$

Provided that X, ε are independent, the marginal density of Y is

$$\int \phi(y-x)\,P_\nu(\mathrm{d}x) = \phi(y)\int e^{-x^2/2}P_\nu(\mathrm{d}x) + \phi(y)\int (e^{xy}-1)e^{-x^2/2}\,P_\nu(\mathrm{d}x),$$

$$= \phi(y)\big(1 - \rho + o(\rho)\big) + \phi(y)\int \big(\cosh(xy)-1\big)e^{-x^2/2}\,P_\nu(\mathrm{d}x),$$

$$= \phi(y)(1-\rho) + \rho\phi(y)\int \big(\cosh(xy)-1\big)e^{-x^2/2}\,H(\mathrm{d}x) + o(\rho),$$

$$= (1-\rho)\phi(y) + \rho\phi(y)\zeta(y) + o(\rho). \tag{5}$$

In the second line, $\int e^{-x^2/2}P_\nu(\mathrm{d}x) = 1 - \rho + o(\rho)$ follows from the definition of the sparsity rate, while symmetry of P_ν permits us to replace the exponential e^{xy} with the cosh function. Since the integrand $(\cosh(xy)-1)e^{-x^2/2}$ is bounded, continuous and $O(x^2)$ at the origin, sparsity allows us replace the P_ν-integral on line 2 with the H-integral in line 3. The zeta function on line 4 is determined by the exceedance measure,

$$\zeta(y) := \int_{\mathbb{R}} (\cosh(xy)-1)e^{-x^2/2}H(\mathrm{d}x).$$

Note that since $\int \phi(y)\zeta(y)\,\mathrm{d}y = 1$ (see, e.g. section 3.2 of [11]), up to first order in the sparsity rate, Y is distributed according to a two-component mixture, ϕ and $\psi = \phi \cdot \zeta$, with corresponding weights $1 - \rho$ and ρ, where ζ is the density ratio of the non-null and null components.

2.4. Convolution for averages

Consider the same set-up as in the previous section where the sample mean observation $\bar{Y} = X + \bar{\varepsilon}$ is the sum of a signal and an independent random

error distributed as $N(0, 1/m)$. For reasons given in section 2.2, the family of signal distributions is assumed to be a sparse scale family with inverse-power exceedance density $H(\mathrm{d}x) = C_d \, \mathrm{d}x / |x|^{d+1}$.

By the same argument given in the preceding section, the marginal density $\int \phi_m(y - x) \, P_\nu(\mathrm{d}x)$ of $\bar{Y}$ at y is equal to

$$
\begin{aligned}
&= \phi_m(y) \int e^{-mx^2/2} P_\nu(\mathrm{d}x) + \phi_m(y) \int (e^{mxy} - 1) e^{-mx^2/2} \, P_\nu(\mathrm{d}x), \\
&= \phi_m(y) \big(1 - \rho_m + o(\rho)\big) + \phi_m(y) \int \big(\cosh(mxy) - 1\big) e^{-mx^2/2} \, P_\nu(\mathrm{d}x), \\
&= \phi_m(y)(1 - \rho_m) + \rho_1 \phi_m(y) \int \big(\cosh(mxy) - 1\big) e^{-mx^2/2} \, H(\mathrm{d}x) + o(\rho), \\
&= (1 - \rho_m) \phi_m(y) + \rho_m \phi_m(y) \zeta(m^{1/2} y) + o(\rho).
\end{aligned}
\tag{6}
$$

Here, $\phi_m(y) = m^{1/2} \phi(m^{1/2} y)$ is the Gaussian density with variance m^{-1}, while the non-Gaussian mixture fraction is

$$
\begin{aligned}
\rho_m &= \int \big(1 - e^{-mx^2/2}\big) P_\nu(\mathrm{d}x), \\
&= \rho \int \big(1 - e^{-mx^2/2}\big) C_d \, \mathrm{d}x / |x|^{d+1} = \rho \, m^{d/2}
\end{aligned}
$$

to first order for the inverse-power measure. Thus, the standardized statistic $m^{1/2} \bar{Y}$ is distributed according to the binary mixture

$$
(1 - \rho_m) \phi(y) + \rho_m \phi(y) \zeta(y) + o(\rho)
\tag{7}
$$

with weight ρ_m on the non-Gaussian component.

The zeta function for the inverse-power exceedance measure is expressible as a power series in y^2 with strictly positive coefficients

$$
\int_{\mathbb{R}} (\cosh(xy) - 1) e^{-x^2/2} \frac{C_d \, \mathrm{d}x}{|x|^{d+1}} = -\sum_{r=1}^{\infty} \frac{(-d)^{\Uparrow r} y^{2r}}{(2r)!} = -\sum_{r=1}^{\infty} \frac{(-d)^{\Uparrow r} y^{2r}}{1^{\Uparrow r} \, 2^r r!}
\tag{8}
$$

[11]. For integer $r \geq 0$, the rising factorial function is the product $\alpha^{\uparrow r} = \alpha(\alpha + 1) \cdots (\alpha + r - 1)$, so that $1^{\uparrow r} = r!$, while $\alpha^{\Uparrow r} = 2^r (\alpha/2)^{\uparrow r}$ is the double-step rising factorial $\alpha^{\Uparrow r} = \alpha(\alpha + 2) \cdots (\alpha + 2(r - 1))$. It follows that $-(-d)^{\Uparrow r}$ is positive for $0 < d < 2$.

2.5. Mixture coefficients

The function $h_\alpha(t) = 1 - (1 - t)^\alpha$ has a series expansion

$$
h_\alpha(t) = -\sum_{r=1}^{\infty} \frac{(-\alpha)^{\uparrow r} t^r}{r!}.
$$

For $0 < \alpha < 1$, the coefficients $\zeta_r = -(-\alpha)^{\uparrow r}/1^{\uparrow r}$ are strictly positive, the tail behaviour is $\zeta_r = O(\Gamma(r - \alpha)/\Gamma(r + 1)) = O(r^{-\alpha - 1})$ for large r, and the series is convergent for $|t| \leq 1$. Since $h_\alpha(1) = 1$, the coefficients determine a probability distribution on the natural numbers. The tail behaviour $\zeta_r = O(r^{-\alpha - 1})$ implies that the moments are not finite. The upper limit $h_1(t) = t$ is an exception implying unit mass at $x = 1$. The distribution with probability generating function h_α arises in the next section in the form of the mixture weights for the non-null mixture component of the marginal density. The correspondence holds when the inverse-power index is $d = 2\alpha$. The probability distribution defined by the coefficients of h_α is further described in appendix A.

2.6. Distribution of the t-ratio

In this section, we derive the distribution of the internally standardized statistic $m^{1/2}\bar{Y}/s$ where s is the sample standard deviation and $\bar{Y}$ is the sample mean among m i.i.d. observations. Let Y denote a random variable with the distribution of $m^{1/2}\bar{Y}$, the two-component mixture (7), and let $s^2 \sim \chi_k^2/k$ be a scalar mean square independent of Y. The sample ratio $T := Y/s$ also has a binary mixture representation with the same weights, where the null component is the t-distribution on $k = m - 1$ degrees of freedom, denoted by f_0, and the non-null component is determined by the zeta function relative to f_0. When the signal distribution has an inverse-power exceedance measure, the Student-t zeta function has an explicit analytic form, stated below and proven in section A.

Theorem 2.2 ([14]). *Let $0 \leq \rho \leq 1$ be given, and suppose that Y is distributed as the mixture*

$$Y \sim (1 - \rho)\phi + \rho\psi, \qquad \psi(y) := \phi(y)\zeta(y),$$

where $\zeta(\cdot)$ is the zeta function (8) for the inverse-power measure H_d with index $d \in (0, 2)$, and Y is independent of $s^2 \sim \chi_k^2/k$. Then the ratio $T = Y/s$ is marginally distributed according to the countably infinite mixture,

$$(1 - \rho)f_0(t) + \rho \sum_{r=1}^{\infty} \zeta_{d,r} f_r(t) \tag{9}$$

where the coefficients $\zeta_{d,r} = -(-d)^{\Uparrow r}/(2^r r!)$ are non-negative and add to one, and the densities f_r are given by

$$f_r(t) = \frac{t^{2r}}{(1 + t^2/k)^{r+1/2+k/2}} \times \frac{\Gamma(1/2)}{k^{r+1/2}\pi^{1/2}B(r + 1/2, k/2)}, \tag{10}$$

where $B(a, b)$ is the Beta function.

It is convenient to write the marginal density (9) as a binary mixture

$$(1 - \rho)f_0(t) + \rho f_0(t)\zeta_k(t), \qquad \zeta_k(t) := \sum_{r=1}^{\infty} \zeta_{d,r} \frac{t^{2r}}{1^{\Uparrow r}} \frac{(1 + k)^{\Uparrow r}}{(k + t^2)^r}. \tag{11}$$

In other words, the density ratio $f_r(t)/f_0(t)$ for $k \geq 2$ is

$$\frac{f_r(t)}{f_0(t)} = \frac{t^{2r}}{1^{\uparrow\uparrow r}} \frac{(1+k)^{\uparrow\uparrow r}}{(k+t^2)^r}. \tag{12}$$

These are the key elementary functions that arise in the definition of the zeta function when the null distribution is Student-t rather than standard normal. Since the series expansion (11) is convergent for all $t \in \mathbb{R}$, the function is well approximated by series truncation; an implementation is provided on the website linked in section 4.

2.7. Probability integral transformation

Let g be the monotone transformation $\mathbb{R} \to \mathbb{R}$ that transforms Student-t to standard normal, i.e., $X \sim f_0$ implies $g(X) \sim N(0,1)$, so g is the probability integral transform that sends t-scores to z-scores. More explicitly, $g(y) = \Phi^{-1}(F_0(y))$, where F_0 is the cumulative distribution function (cdf) of the t-distribution with k degrees of freedom, and Φ is the cdf of the standard normal distribution.

Proposition 2.1. *Let $d \in (0, 2)$, and suppose T is distributed according to the t-mixture* (9). *Then the transformed variable $Z = gT$ is distributed as a mixture with density*

$$\phi(z) \left(1 - \rho + \rho\zeta_k(g^{-1}z)\right), \tag{13}$$

where $g^{-1}z$ is the t-score. By contrast, if Y is distributed as in Theorem 2.2, it has density

$$\phi(y) \left(1 - \rho + \rho\zeta_\infty(y)\right),$$

where $\zeta_\infty(y) := \lim_{k \to \infty} \zeta_k(y)$.

Suppose our data consist of independent replicates (Y_i, s_i^2) for $i = 1, \ldots, n$ from the model (1), with a different signal X_i for each site. Classical procedures in the multiple testing literature (e.g. [13], [1]) typically take as input a list of p-values. This reduction of the data, obtained by the probability integral transform of the t-scores, subsequently ignores the degree of freedom parameter k in the calculation of false discovery rates. As expression (13) indicates, the density of the non-null component depends on $k = m - 1$, whereas running the p-values through a general purpose multiple testing procedure such as Lindsey's method ([9], [10], [5]) or the BH procedure effectively assumes that the convolution occurs in the space of z-scores. The formula obtained in Theorem 2.2 for the zeta function relative to the Student-t null facilitates a numerical comparison between these two approaches.

3. Numerical comparison

Although there are close similarities, the probability-integral transformed t-mixture in (13) is not the same as the standard Gaussian mixture

$$(1 - \rho)\phi(z) + \rho\phi(z)\zeta_\infty(z).$$

The zeta-function for the transformed mixture is $\zeta_k(g^{-1}z)$, which is close to, but not the same as $\zeta_\infty(z)$. The goal of this section is to demonstrate the effect of this difference on false discovery rate calculations numerically and on the HIV dataset of [5].

The two-groups model is a description of the data generating process that posits a latent variable indicating the component from which the observation arose,

$$H \sim \text{Bernoulli}(\rho)$$

$$Z \mid H \sim \begin{cases} \phi & \text{if } H = 0 \\ \psi & \text{if } H = 1, \end{cases}$$

where $\psi(z) = \phi(z)\zeta(z)$ is an alternative density satisfying $\psi(0) = 0$. A quantity of interest is the local false discovery rate ([6]), defined as the conditional probability that the latent variable H is zero,

$$\text{lfdr}(z) = \mathbb{P}(H = 0 \mid Z = z) = \frac{(1 - \rho)\phi(z)}{(1 - \rho)\phi(z) + \rho\psi(z)} \, . \tag{14}$$

The zeta function together with the sparsity rate determine the conditional odds ratio,

$$\frac{1 - \text{lfdr}(z)}{\text{lfdr}(z)} = \frac{\rho}{1 - \rho} \times \zeta(z) \, .$$

For small degrees of freedom (≤ 10), the conditional odds that a signal is active can be diminished by roughly 20% in the region of interest (e.g. $3 \leq |z| \leq 10$). The ratios $\zeta_\infty(z)/\zeta_k(t)$, with $z = g(t)$, are shown below for $k = 10$ and three values of d, the parameter governing the tail behavior of the signal distribution:

TABLE 1
Ratios $\zeta_\infty(z)/\zeta_{10}(t)$ for $k = 10$.

t	z	$d = 0.5$	$d = 1.0$	$d = 1.5$
3.0	2.47	0.84	0.94	1.05
4.0	3.02	0.79	0.93	1.10
5.0	3.46	0.73	0.90	1.12
7.0	4.12	0.65	0.85	1.12
10.0	4.80	0.57	0.82	1.18
15.0	5.51	0.52	0.85	1.37

It is apparent that the ratio $\zeta_\infty(z)/\zeta_{10}(t)$ may be appreciably less than one for heavy-tailed signals with $d < 1$, or appreciably greater than one for shorter-tailed signals with $d > 1$. Also, the ratio is not monotone as a function of the argument.

The region of most interest for comparison is typically $3 \leq |z| \leq 6$. Ordinarily, $k = 100$ and $k = \infty$ are effectively equivalent for most statistical purposes. In sparsity calculations, however, the dependence on the degrees of freedom is far from negligible in the region of interest, even for k above 100. For $d = 1$, the zeta-ratios $\zeta_\infty(z)/\zeta_{100}(z)$ at $z = 3, 4, 5, 6$ are 1.14, 1.62, 3.37 and 11.68 respectively.

3.1. *Comparison on HIV data*

The HIV data, taken from Chapter 6 of [5], is a case-control study with gene-activity levels measured at 7680 genomic sites. For each site we compute a difference between the average gene expression levels for case and control, consisting of 4 HIV-positive and 4 HIV-negative individuals, respectively. The histogram of resulting differences is asymmetric, and although the methods we demonstrate in this section here assume a symmetric signal distribution, we still find it worthwhile to make the comparison on this dataset, as it has a small degree of freedom parameter ($k = 6$) for the pooled variance estimate.

In the analysis by [5], the goal of the study was to identify a small subset of the genes that are potentially relevant towards understanding HIV. Each difference was divided by a pooled standard error to obtain a t-score, and these were transformed into z-scores via $Z_i = \bar{\Phi}^{-1}(1 - F_0(T_i))$, where $\bar{\Phi} = 1 - \Phi$ is the complement of the standard normal cdf and $\bar{\Phi}^{-1}$ is its quantile function. After this pre-processing step, the (Z_i) were viewed as independent samples from a two-groups model with standard Gaussian null component,

$$H_i \overset{\text{iid}}{\sim} \text{Bernoulli}(\rho)$$
$$Z_i \mid H_i \sim (1 - H_i)\phi + H_i\psi, \quad \text{independently.}$$

If ψ is the non-null component defined by the inverse-power exceedance with index $d \in (0, 2)$, the maximized log likelihood relative to the null model is 48.23, which occurs at

$$\hat{\rho}_z = 0.0059, \quad \hat{d}_z = 1.09 \,.$$

In formula (14), these values yield an estimate $\hat{\ell}_{z,i}$ of the lfdr for each site $i = 1, \ldots, 7680$. Assuming the null component is instead Student-t, i.e., each T_i comes from the two-groups model (9), the maximized log likelihood is 56.13, which occurs at

$$\hat{\rho}_t = 0.0045, \quad \hat{d}_t = 0.60 \,.$$

Combining these estimates with the formula for the zeta function when the null is Student-t gives an analogous set of local fdr estimates

$$\hat{\ell}_{t,i} := \frac{1 - \hat{\rho}_t}{1 - \hat{\rho}_t + \hat{\rho}_t \zeta_k(T_i)}, \quad i = 1, \ldots, 7680 \,.$$

ζ_k can be computed quickly and accurately by truncating the infinite series representation (11). The resulting local fdr estimates for the sites with the largest (in absolute value) test statistics are displayed in Table 2. The BH procedure at level $\alpha = 0.1$ yields 16 rejections, and the $\hat{\ell}_{t,i}$ values among these range from 0.00 to 0.42, with an average of 0.108. Within the BH (0.1) set, the $\hat{\ell}_{z,i}$ values range from 0.00 to 0.51, with an average of 0.144.

We also compare the lfdr estimates with $d = 1$ fixed, i.e., plugging $d = 1$ into the formula (14) for both t and z scores instead of a maximum likelihood

TABLE 2
Estimated local fdr values for the top 6 test statistics.

Z_i	T_i	$\hat{\ell}_{z,i} \times 10^4$	$\hat{\ell}_{t,i} \times 10^4$	$\hat{\ell}_{z,i}/\hat{\ell}_{t,i}$
-5.91	-52.16	1.54	0.77	2.00
-5.74	-43.89	3.93	1.95	2.01
-5.60	-38.36	8.12	4.03	2.02
-5.53	-35.86	11.65	5.78	2.01
-5.40	-31.57	22.94	11.44	2.01
-5.13	-24.74	82.10	41.86	1.96

estimate $\hat{d}_z$ or $\hat{d}_t$, which differ non-negligibly for the HIV data. The resulting estimates of lfdr are shown in Table 3. In other words, the factor of 2 in the last column of Table 2 could arise in part from the difference between $\hat{d}_z$ and $\hat{d}_t$, and this effect is controlled for by fixing $d = 1$ in the computation for $\hat{\ell}_z$ and $\hat{\ell}_t$.

The maximum likelihood estimates of the sparsity rates are $\hat{\rho}_z = 0.0053$ and $\hat{\rho}_t = 0.0078$, while the maximized log likelihoods are 48.13 and 52.99 relative to the null ($\rho = 0$). Since the estimated power index for the z-scores is close to 1 ($\hat{d}_z = 1.09$), the resulting lfdr estimates are nearly the same as the ones in Table 2 even when $d = 1$ is held fixed rather than estimated.

TABLE 3
Estimated local fdr values for the top 6 test statistics with $d = 1$.

Z_i	T_i	$\hat{\ell}_{z,i} \times 10^4$	$\hat{\ell}_{t,i} \times 10^4$	$\hat{\ell}_{z,i}/\hat{\ell}_{t,i}$
-5.91	-52.16	1.52	1.43	1.06
-5.74	-43.89	3.88	3.38	1.15
-5.60	-38.36	8.02	6.60	1.22
-5.53	-35.86	11.53	9.22	1.25
-5.40	-31.57	22.78	17.32	1.32
-5.13	-24.74	82.01	57.30	1.43

The values of $\hat{\ell}_{t,i}$ among the 16 BH($\alpha = 0.1$) rejections now range from 0.00 to 0.39 with an average value of 0.104, whereas the values $\hat{\ell}_{z,i}$ among the rejections range from 0.00 to 0.52 with an average value of 0.147.

4. Discussion

In this paper, we derived a mixture representation of the non-null density of a t-ratio when the signal is sparse, and illustrated the maximum-likelihood procedure for estimating the local false discovery rate when the signal distribution has an inverse-power exceedance measure. The formula for the density is determined by the Student-t zeta function and the sparsity rate ρ, and depends explicitly on the degrees of freedom parameter k. When $k \to \infty$, the formula recovers the Bayes factor for the standard Gaussian mixture in the sparse setting. For small k, we have demonstrated differences in the deduced local false discovery rates, both numerically in a typical region of interest, and on a dataset of

high-throughput gene expression levels of HIV patients. Although neither model accommodates the asymmetry that is present in the HIV data, the Student-t-model fits appreciably better than the transformed z-model in terms of the log-likelihood, and the average fitted lfdr-value within typical BH_α subsets is a reasonably close match with α.

The analysis presented here is agnostic to the assumption of equal variances across genes. If there were reason to believe that the variance was constant across sites, we could pool the estimates, yielding a scaled χ^2 variable with essentially infinite degrees of freedom. Our analysis also assumes that the sparsity rate is the same at every site, and the joint distribution of (X_i, Y_i) depends on (ρ, d, σ_i). But the local false discovery formula (14) is agnostic on the question of independence from one site to another.

Reproducibility. The R code and data used to generate Tables 2 and 3 can be found at the following Github link: https://github.com/dan-xiang/dan-xiang.github.io/tree/master/t-statistics-paper

Appendix A: Proofs

Proof of Example 2.1. Let $P_\sigma(x) = p_\sigma(x)\mathrm{d}x$ denote the Student-t distribution on d degrees of freedom with standard density:

$$p_1(x) = \frac{\Gamma(\frac{d+1}{2})}{\sqrt{d\pi}\Gamma(d/2)}\left(1 + \frac{x^2}{d}\right)^{-\frac{(d+1)}{2}}$$

It suffices to show

$$\sigma^{-d}\frac{C_d\sqrt{\pi}\Gamma(d/2)}{d^{d/2}\Gamma(\frac{d+1}{2})}p_\sigma(x) \to \frac{C_d}{|x|^{d+1}}.$$

for each $x \neq 0$. The left hand side is equal to

$$\rho^{-1}p_\sigma(x) = \sigma^{-d} \cdot \frac{C_d\sqrt{\pi}\Gamma(d/2)}{d^{d/2}\Gamma(\frac{d+1}{2})} \cdot \frac{1}{\sigma}\frac{\Gamma(\frac{d+1}{2})}{\sqrt{d\pi}\Gamma(d/2)}\left(1 + \frac{(x/\sigma)^2}{d}\right)^{-\frac{(d+1)}{2}}$$

$$= \sigma^{-(d+1)} \cdot \frac{C_d}{d^{d/2}\sqrt{d}} \cdot \left(\sigma^2 + \frac{x^2}{d}\right)^{-\frac{(d+1)}{2}}\sigma^{d+1}$$

$$\to \frac{C_d}{d^{(d+1)/2}} \cdot \frac{d^{\frac{d+1}{2}}}{|x|^{d+1}} = \frac{C_d}{|x|^{d+1}}.$$

$\square$

Proof of Theorem 2.1. Let $P_\sigma(\mathrm{d}x) = p_\sigma(x)\mathrm{d}x$ denote a sparse scale family with standard density p_1, sparsity rate $\rho(\sigma)$, and exceedance measure $H(\mathrm{d}x) = h(x)\mathrm{d}x$. Then as $\sigma \to 0$

$$p_\sigma(x) = \frac{p_1(x/\sigma)}{\sigma} \sim \rho(\sigma)h(x)$$

for any $x \neq 0$. Taking $x = 1$, the above equivalence gives

$$p_1(u) \sim h(1)\rho(u^{-1})|u|^{-1} \text{ as } |u| \to \infty,$$

which implies that for any $x > 0$, as $\sigma \to 0$

$$h(x) \sim \frac{p_1(x/\sigma)}{\sigma\rho(\sigma)} \sim \frac{h(1)\rho(\sigma/x) \cdot (\sigma/x)}{\sigma\rho(\sigma)} = h(1) \cdot \frac{\rho(\sigma/x)}{x\rho(\sigma)}.$$

Equivalently, as $\sigma \to 0$,

$$\frac{\rho(\sigma/x)}{\rho(\sigma)} \to \frac{xh(x)}{h(1)}.$$

Let $\gamma(u) := \frac{1}{\rho(u^{-1})}$. Then the above convergence implies that for any $x > 0$,

$$\lim_{u \to \infty} \frac{\gamma(xu)}{\gamma(u)} = \frac{h(1)}{xh(x)}. \tag{15}$$

This condition implies $h(x) = h(1)x^{-d-1}$ for some $d \in \mathbb{R}$ (see Lemma 1 in Chapter 8 of [7]). Therefore $\gamma(u) = u^d L(u)$ for some slowly varying function $L : (0, \infty) \to (0, \infty)$. Indeed, let $L(u) := \gamma(u)/u^d$. Then for any $a > 0$, (15) implies as $u \to \infty$

$$\frac{L(au)}{L(u)} = \frac{\gamma(au)u^d}{\gamma(u)(au)^d} = \frac{\gamma(au)}{\gamma(u)a^d} \to \frac{h(1)}{ah(a)}a^{-d} = \frac{h(1)}{a \cdot h(1)a^{-d-1}}a^{-d} = 1.$$

In particular, this implies $\rho(\sigma) = \sigma^d/L(\sigma^{-1})$. Further, we deduce $d < 2$ since sparsity implies $h(x)$ must be integrable against functions satisfying $w(x) = O(x^2)$ as $x \to 0$. □

Properties of the distribution with probability generating function h_α (section 2.5). For X distributed according h_α, the reciprocal moment is the harmonic number

$$E(1/X) = \int_0^1 t^{-1} h_\alpha(t)\, \mathrm{d}t = \psi(\alpha + 1) - \psi(1),$$

where ψ is the derivative of the log gamma function. The inverse factorial moments are

$$E\left(\frac{r!}{(X+1)^{\uparrow r}}\right) = \frac{\alpha}{\alpha + r}.$$

In the special case $\alpha = 1/2$, the probabilities are $\zeta_r = C_{r-1}/2^{2r-1}$, where $C_r = (2r)!/(r!(r+1)!)$ is the rth Catalan number. Since $h_\alpha^2(t) = 2h_\alpha(t) - h_{2\alpha}(t)$ and $h_1(t) = t$, the distribution $h_{1/2}$ has the peculiar property that the two-fold convolution is equal to the conditional distribution given $X > 1$. It is the unique distribution on the natural numbers having this property. □

Proof of Theorem 2.2. It follows from Lemma A.1 that ψ is a mixture

$$\psi(y) = \sum_{r=1}^{\infty} \zeta_{d,r} \cdot g_r(y), \quad g_r(y) := \frac{y^{2r}\phi(y)}{1^{\uparrow\uparrow r}}.$$

The fact that g_r is a valid probability density follows from $\int y^{2r}\phi(y)\,\mathrm{d}y = 1^{\uparrow\uparrow r}$, the formula for a Gaussian moment of even degree. If $X_r \sim g_r$ and $S^2 \sim \chi_k^2$, then by Lemma A.2, the density of $T_r = X_r\sqrt{k/S^2}$ is given by

$$f_r(t) = \frac{t^{2r}}{(1+t^2/k)^{r+1/2+k/2}} \times \frac{\Gamma(1/2)}{k^{r+1/2}\pi^{1/2}B(r+1/2,k/2)},$$

so the marginal distribution of $T = Y\sqrt{k/S^2}$ can be written

$$(1-\rho)f_0(t) + \rho\sum_{r=1}^{\infty} \zeta_{d,r}f_r(t).$$

$\square$

Proof of Proposition 2.1. The density of Z is

$$\left[(1-\rho)f_0(g^{-1}z) + \rho f_0(g^{-1}z)\zeta_k(g^{-1}z)\right](g^{-1})'(z).$$

By the inverse function theorem,

$$(g^{-1})'(z) = \frac{1}{g'(g^{-1}z)} = \frac{\phi(z)}{f_0(g^{-1}z)},$$

which implies the first formula. For the second formula, note that by Lemma A.1, the density of Y can be written $(1-\rho)\phi(y) + \rho\phi(y)\zeta(y)$, where

$$\zeta(y) = \sum_{r=1}^{\infty} \zeta_{d,r}\frac{y^{2r}}{1^{\uparrow\uparrow r}}, \quad \zeta_{d,r} = -\frac{(-d)^{\uparrow\uparrow r}}{2^r r!}.$$

Now since $\lim_{k\to\infty} \frac{(1+k)^{\uparrow\uparrow r}}{(k+y^2)^r} = 1$ for fixed $r \in \mathbb{N}$ and $y \in \mathbb{R}$,

$$\zeta_k(y) = \sum_{r=1}^{\infty} \zeta_{d,r}\frac{y^{2r}}{1^{\uparrow\uparrow r}}\frac{(1+k)^{\uparrow\uparrow r}}{(k+y^2)^r} \to \sum_{r=1}^{\infty} \zeta_{d,r}\frac{y^{2r}}{1^{\uparrow\uparrow r}} = \zeta(y) \quad \text{as } k \to \infty,$$

since the sequence $\frac{(1+k)^{\uparrow\uparrow r}}{(k+y^2)^r}$ is eventually monotone in k, and the series is convergent for each fixed k. $\square$

Lemma A.1. *Suppose Y distributed*

$$Y \sim (1-\rho)\phi + \rho\psi, \quad \psi(y) := \phi(y)\zeta(y)$$

$$\zeta(y) = \int_{\mathbb{R}\setminus\{0\}} (\cosh(xy) - 1)e^{-x^2/2}H_d(\mathrm{d}x)$$

where H_d is the unit inverse-power exceedance defined in (2.1) with index $d \in (0, 2)$. Then

$$\zeta(y) = \sum_{r=1}^{\infty} \zeta_{d,r} \frac{y^{2r}}{1^{\uparrow\uparrow r}}, \quad \zeta_{d,r} := -\frac{(-d)^{\uparrow\uparrow r}}{2^r r!}.$$

Proof. By definition, the zeta function for Y is

$$\begin{aligned}
\zeta(y) &= \int_{\mathbb{R}\backslash\{0\}} (\cosh(xy) - 1) e^{-x^2/2} H_d(\mathrm{d}x) \\
&= \int_{\mathbb{R}\backslash\{0\}} \sum_{r=1}^{\infty} \frac{(xy)^{2r}}{(2r)!} e^{-x^2/2} H_d(\mathrm{d}x) \\
&= \sum_{r=1}^{\infty} \frac{y^{2r}}{1^{\uparrow\uparrow r}} \cdot \frac{1^{\uparrow\uparrow r}}{(2r)!} \int_{\mathbb{R}\backslash\{0\}} x^{2r} e^{-x^2/2} H_d(\mathrm{d}x).
\end{aligned}$$

It remains to show that

$$\frac{1^{\uparrow\uparrow r}}{(2r)!} \int_{\mathbb{R}\backslash\{0\}} x^{2r} e^{-x^2/2} H_d(\mathrm{d}x) = -\frac{(-d)^{\uparrow\uparrow r}}{2^r r!}. \tag{16}$$

The left hand side evaluates to

$$\begin{aligned}
\frac{1^{\uparrow\uparrow r}}{(2r)!} \int_{\mathbb{R}\backslash\{0\}} x^{2r} e^{-x^2/2} H_d(\mathrm{d}x) &= \frac{1^{\uparrow\uparrow r}}{(2r)!} \int_{\mathbb{R}\backslash\{0\}} x^{2r} e^{-x^2/2} \frac{d2^{d/2-1}}{\Gamma(1-d/2)} \frac{\mathrm{d}x}{|x|^{1+d}} \\
&= \frac{2^{d/2}}{-\Gamma(-d/2)2^r r!} \int_{\mathbb{R}\backslash\{0\}} |x|^{2r-1-d} e^{-x^2/2} \, \mathrm{d}x,
\end{aligned}$$

using the definition $\Gamma(z+1) = z\Gamma(z)$. Now using that $\mathbb{E}(|Z|^p) = \frac{2^{p/2}\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}$ for $p > -1$ and $Z \sim N(0, 1)$, the above becomes

$$\begin{aligned}
&= \frac{2^{d/2}\sqrt{2\pi}}{-\Gamma(-d/2)2^r r!} \cdot \frac{2^{\frac{2r-1-d}{2}} \Gamma\left(\frac{2r-1-d+1}{2}\right)}{\sqrt{\pi}} \\
&= \frac{\Gamma(r-d/2)}{-\Gamma(-d/2)r!} = \frac{d(2-d)\cdots(2r-2-d)}{2 \cdot 4 \cdots (2r)} = -\frac{(-d)^{\uparrow\uparrow r}}{2^r r!}
\end{aligned}$$

which is monotone decreasing at rate $O(r^{-1-d/2})$ for large r. $\qquad\square$

Lemma A.2. *Suppose $r \geq 1$ and $X_1 \sim g_r$ where $g_r(x) = \frac{x^{2r}\phi(x)}{1^{\uparrow\uparrow r}}$ and $X_2^2 \sim \chi_k^2$ independently. Then the density of $T_1 := X_1\sqrt{k/X_2^2}$ is given by*

$$f_r(t) = \frac{\Gamma(1/2)}{k^{r+1/2}\pi^{1/2}B(r+1/2, k/2)} \cdot \frac{t^{2r}}{(1+t^2/k)^{r+1/2+k/2}}.$$

Proof. Independence implies that the joint density of X_1, X_2 on $\mathbb{R}^d \times \mathbb{R}^+$ is the product

$$x_1^{2r} e^{-x_1^2/2} x_2^{k/2-1} e^{-x_2/2} \times \frac{1}{(2\pi)^{1/2} 1^{\Uparrow r}} \frac{1}{2^{k/2}\Gamma(k/2)}\,.$$

The transformation $(x_1, x_2) \mapsto (t_1, t_2) = (x_1\sqrt{k/x_2}, x_2)$ has Jacobian $\sqrt{t_2/k}$, so the joint distribution of the transformed variables is

$$t_1^{2r} (t_2/k)^r e^{-t_2(t_1^2/k+1)/2}\, t_2^{k/2-1} \sqrt{t_2/k} \times \frac{1}{(2\pi)^{1/2} 1^{\Uparrow r} 2^{k/2}\Gamma(k/2)}$$

$$= \frac{1}{k^{r+1/2}(2\pi)^{1/2} 1^{\Uparrow r} 2^{k/2}\Gamma(k/2)} t_1^{2r} \cdot t_2^{r+1/2+k/2-1} e^{-t_2(t_1^2/k+1)/2}\,.$$

We recognize the pdf of the Gamma $\left(r + 1/2 + k/2, (1 + t_1^2/k)/2\right)$ distribution in the above expression, and integrate over $t_2 \in \mathbb{R}^+$ to obtain the marginal density of T_1 at $t_1 \in \mathbb{R}$:

$$f_r(t_1) = \frac{t_1^{2r}}{(1 + t_1^2/k)^{r+1/2+k/2}} \times \frac{\Gamma(r+1/2+k/2)\, 2^{r+1/2+k/2}}{k^{r+1/2}(2\pi)^{1/2} 1^{\Uparrow r}\, 2^{k/2}\Gamma(k/2)}$$

$$= \frac{t_1^{2r}}{(1 + t_1^2/k)^{r+1/2+k/2}} \times \frac{\Gamma(1/2)}{k^{r+1/2}\pi^{1/2} B(r + 1/2, k/2)}\,,$$

where $B(\cdot, \cdot)$ is the beta function. Note that $f_0(\cdot)$ is the Student-t density on k degrees of freedom on $\mathbb{R}$. $\qquad\square$

References

[1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society: series B (Methodological)* **57**(1): 289–300. MR1325392

[2] Carvalho, C. M., Polson, N. G. and Scott, J. G. (2010). The horseshoe estimator for sparse signals, *Biometrika* **97**(2): 465–480. MR2650751

[3] Dudoit, S., Shaffer, J. P. and Boldrick, J. C. (2003). Multiple hypothesis testing in microarray experiments, *Statistical Science* **18**(1): 71–103. MR1997066

[4] Efron, B. (2008). Microarrays, empirical bayes and the two-groups model, *Statistical Science* **23**: 1–22. MR2431866

[5] Efron, B. (2012). *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, Vol. 1, Cambridge University Press. MR2724758

[6] Efron, B., Tibshirani, R., Storey, J. D. and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment, *Journal of the American Statistical Association* **96**(456): 1151–1160. MR1946571

[7] Feller, W. (1991). *An introduction to probability theory and its applications, Volume 2*, Vol. 81, John Wiley & Sons. MR0038583

[8] Ignatiadis, N. and Sen, B. (2023). Empirical partially bayes multiple testing and compound χ^2 decisions, *arXiv preprint arXiv:2303.02887*.

[9] Lindsey, J. (1974a). Comparison of probability distributions, *Journal of the Royal Statistical Society: Series B (Methodological)* **36**(1): 38–47. MR0362643

[10] Lindsey, J. (1974b). Construction and comparison of statistical models, *Journal of the Royal Statistical Society: Series B (Methodological)* **36**(3): 418–425. MR0365794

[11] McCullagh, P. and Polson, N. G. (2018). Statistical sparsity, *Biometrika* **105**(4): 797–814. MR3877866

[12] Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression, *Journal of the American Statistical Association* **83**(404): 1023–1032. MR0997578

[13] Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance, *Biometrika* **73**(3): 751–754. MR0897872

[14] Tresoldi, M. F. (2021). *Sparsity*, PhD thesis, The University of Chicago. MR4336693