Supplementary Information for Selection Corrected Statistical Inference for Region Detection with High-throughput Assays

Yuval Benjamini

Department of Statistics, Hebrew University of Jerusalem, Israel
and
Jonathan Taylor
Department of Statistics, Stanford University
and
Rafael A. Irizarry
Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute
and Department of Biostatistics, Harvard University

July 5, 2018

Abstract

Supplementary information for the companion manuscript Selection Corrected Statistical Inference for Region Detection with High-throughput Assays. The supplementary includes a section on the accelerated sampling algorithm, proofs for the lemmas, and additional information about the simulation and the data experiments.

Contents

1 Accelerated sampling

Recall that we parameterize the truncated multivariate normal (TMN) family with a single mean parameter θ that linearly determines the mean vector. The distributions corresponding to different values of θ form a single parameter exponential family; this implies that importance weighting can efficiently convert a sample from f_{θ_0} into a sample from f_{θ_1}. We detail here the algorithm, with notations following Owen (2013). This an expansion of the ideas described in the appendix of Fithian et al. (2014).

1.1 Importance sampling for the truncated multivariate normal

Suppose we have a Monte Carlo sample $\mathbf{z}_1, ...\mathbf{z}_N \sim f_{\theta_0}$ and we would like to estimate of $\mathbf{E}[g(Z)]$ for $Z \sim f_{\theta_1}$. Then the importance sampling estimate of $\mathbf{E}_{\theta_1}[g(Z)]$ is

$$\hat{\mathbf{E}}_{\theta_1}[g(Z)] = \sum w_i(\mathbf{z}_i) g(\mathbf{z}_i)$$

where

$$w_i \propto \frac{f_{\theta_1}(\mathbf{z}_i)}{f_{\theta_0}(\mathbf{z}_i)}, \qquad \sum w_i = 1.$$

The importance estimator is unbiased. With a careful choice of θ_0 it may enjoy lower variance per sample-size compared to Monte Carlo estimates from f_{θ_1}. Nevertheless, for our application the primary gain is the ability to invert tests for numerous values of θ using a single sample.

The exponential tilting principle (Siegmund, 1976) recognizes that for single-parameter exponential families, the importance weights $w_1, ..., w_n$ can be calculated without explicitly calculating the normalizing constant for the destination density f_{θ_1}. We develop here the explicit form for the TMN densities parameterized by a linear mean shift.

The TMN density for a region of size $b - a + 3$ is written in full as:

$$f_{\theta;\mathbf{s}} = f_{Z\ |A_{a:b};\theta,\mathbf{s},\Sigma}(\mathbf{z}) = \frac{\exp\{(\mathbf{z} - \mu(\theta))'\Sigma^{-1}(\mathbf{z} - \mu(\theta))\}}{\int_{A_{a:b}} \exp\{(\mathbf{u} - \mu(\theta))'\Sigma^{-1}(\mathbf{u} - \mu(\theta))\} d\mathbf{u}} 1(\mathbf{z} \in A_{a:b}), \tag{1}$$

with the mean vector parameterized linearly in θ

$$\mu(\theta) = \mu_0 + \theta \cdot \mathbf{s}. \tag{2}$$

In order to accommodate the external constraints (Section 4 of the main paper), we have

$$\mu_0 = (Z^{obs}_{a-1}, 0, 0, ..., 0, Z^{obs}_{b+1})$$

and a profile $\mathbf{s}$ such as $\mathbf{s} = \frac{1}{a-b+1}(0, 1, 1, ..., 1, 0)$.

We can expand this density to exponential family form:

$$\begin{aligned} f_{Z \mid A_{a:b};\Sigma}(\mathbf{z};\theta) &= \exp\left\{\mathbf{z}\Sigma^{-1}\mu(\theta) + h(\mathbf{z}) - g(\theta)\right\} 1_{A_{a:b}}(\mathbf{z}), \\ &= \exp\left\{\mathbf{z}\Sigma^{-1}\mu_0 + \theta\mathbf{z}\Sigma^{-1}\mathbf{s} + h(\mathbf{z}) - g(\theta)\right\} 1_{A_{a:b}(\mathbf{z})} \end{aligned}$$

where $h(\mathbf{z}) = -\frac{1}{2}(\mathbf{z}\Sigma^{-1}\mathbf{z})$ does not depend on θ, and $g(\theta) = -\frac{1}{2}\mu(\theta)'\Sigma^{-1}\mu(\theta) - \int_{A_{a:b}} \phi(\mathbf{u}, \mu(\theta), \Sigma)d\mathbf{u}$ is the normalizing constant. Therefore, the likelihood ratio for an example $\mathbf{z} \in A_{a:b}$ depends on the covariance-corrected shape

$$\frac{f_{\theta_1}}{f_{\theta_0}}(\mathbf{z}) = \tilde{w}_i(\mathbf{z}) = d_{\theta_0\theta_1} \exp\left\{(\theta_1 - \theta_0)\mathbf{z}'\tilde{\mathbf{s}}\right\}, \quad \tilde{\mathbf{s}} = \Sigma^{-1}\mathbf{s}$$

where the factor $d_{\theta_0\theta_1} = \exp\{g(\theta_0) - g(\theta_1)\}$ does not depend on $\mathbf{z}$. Instead of computing $d_{\theta_0\theta_1}$ the weights for a given Monte Carlo sample $\tilde{w}_1(\mathbf{z}_1), ..., \tilde{w}_n(\mathbf{z}_n)$ can be normalized

$$w_i(\mathbf{z}_i) = \frac{\tilde{w}_i(\mathbf{z}_i)}{\sum_{j=1}^n \tilde{w}_j(\mathbf{z}_j)} = \frac{\exp\left\{(\theta_1 - \theta_0)\mathbf{z}_i'\tilde{\mathbf{s}}\right\}}{\sum_j \exp\left\{(\theta_1 - \theta_0)\mathbf{z}_j'\tilde{\mathbf{s}}\right\}}$$

to meet both conditions of empirical importance sampling weights.

The main speed-up in tilting arises from the ability to use a single sample from f_{θ_0} to identify the acceptance region for any θ. This is particularly useful because our method for estimating confidence intervals requires building many individual tests for a dense grid of θ values. The main computational load comes from sampling the TMN under the constraint; tilting allows us to perform this task only once, for a convenient value θ_0, and to extract all tests from this sample.

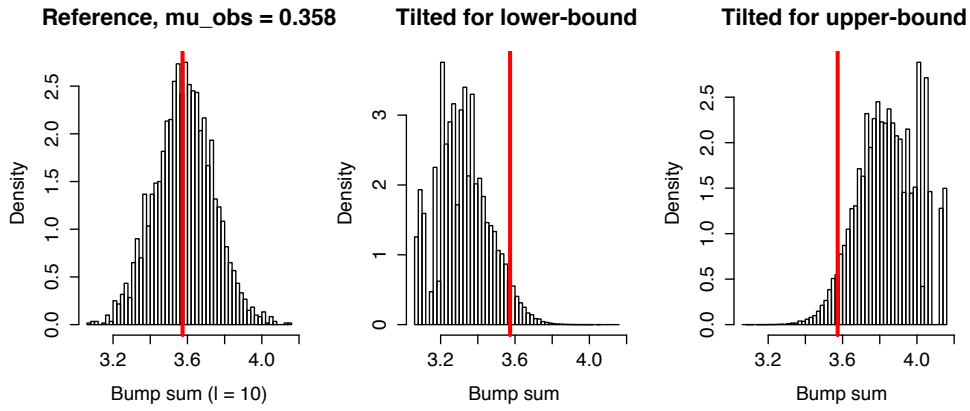

Figure 1: Example of how to extract confidence intervals from a reference density. We generate a sample from a reference density (and mean parameter θ) for which the observed statistic $t_{obs} = 0.358$ (red vertical line) is in the bulk of the distribution. The sample can then be tilted - reweighed - to generate other distributions within the same parametric family. The tilted distribution giving the lower bound of the confidence interval is plotted in the center. This is the most left-tilted distribution for which $P(T > t_{obs}) \geq \alpha/2$. The distribution for the upper-bound is plotted on the right.

1.2 Estimating p-values, quantile functions and intervals

Estimating the p-value

For an observe value t_{obs}, the one-sided p-value is defined as

$$\text{p-value} = \mathbf{P}_{\theta=0}[t(Z) \geq t_{obs}]. \tag{3}$$

To estimate the p-value from a sample of $\mathbf{z}_1, ..., \mathbf{z}_N \sim f_{\theta_0}$, we tilt the sample for $\theta_1 = 0$, and plug in Equation (3)

$$\widehat{\text{p-value}} = \sum_i \frac{\exp\{-\theta_0 \mathbf{z}_i' \tilde{\mathbf{s}}\}}{\sum_{j=1}^n \exp\{-\theta_0 \mathbf{z}_j' \tilde{\mathbf{s}}\}} 1(t(Z) \geq t_{obs}).$$

Computing an interval

The interval is constructed by inverting a family of tests for a grid of parameters $\{\theta_m\}_{m=1}$. For a pre-specified θ_0 and sample $\mathbf{z}_1, ..., \mathbf{z}_N \sim f_{\theta_0}$, we can explicitly construct two sided $1-\alpha$ level tests for each value θ_i by estimating the quantile functions

$$\hat{q}_{\alpha/2;\theta_m} = \text{argmax}_q\{q \in \mathbf{R} : \sum_{i:t(\mathbf{z}_i)>q} \frac{\exp\{(\theta_m - \theta_0)\mathbf{z}_i' \tilde{\mathbf{s}}\}}{\sum_{j=1}^n \exp\{(\theta_m - \theta_0)\mathbf{z}_j' \tilde{\mathbf{s}}\}} \leq \alpha/2\},$$

$$\hat{q}_{1-\alpha/2;\theta_m} = \text{argmin}_q\{q \in \mathbf{R} : \sum_{i:t(\mathbf{z}_i)>q} \frac{\exp\{(\theta_m - \theta_0)\mathbf{z}_i' \tilde{\mathbf{s}}\}}{\sum_{j=1}^n \exp\{(\theta_m - \theta_0)\mathbf{z}_j' \tilde{\mathbf{s}}\}} \geq 1-\alpha/2\}.$$

The two sided test for $\theta = \theta_m$ is therefore:

$$\phi^{(2)}_{\theta_m,\alpha}(Z) = 1(t(Z) \in [\hat{q}_{\alpha/2;\theta_m}, \hat{q}_{\alpha/2;\theta_m}]),$$

and the $1-\alpha$ confidence interval can be constructed as the interval that includes only non-rejected θ_m's:

$$I_{1-\alpha}(Z) = range(\{\theta_m : \phi^{(2)}_{\theta_m,\alpha}(Z) = 1\}).$$

Note that the estimates of quantiles might not be monotone in θ_m. Lemmas 1 and 2 of Section 4.2.1 of the main paper discuss conditions that assure that the true quantile functions are monotone in θ_m. However, even under these conditions, the observed ones might not be. The reason is that for a specific sample $\{\mathbf{z}_1, ..., \mathbf{z}_N\}$ the statistic $t(\mathbf{z})$ and the weight $\mathbf{z}'\tilde{\mathbf{s}}$ may not produce same ranking. They coincide exactly under the conditions of Lemma 1, when $t(\mathbf{z}) = \mathbf{z}'\Sigma^{-1}\mathbf{s}$ is the sufficient statistic.

1.3 Additional considerations

Choosing θ_0

Heuristically, we prefer a θ_0 for which t_{obs} is a likely outcome, with enough samples on both sides of t_{obs}. In the extreme case, to identify any p-value in $(0, 1)$ via tilting requires that $0 \leq \hat{\mathbf{P}}_{\theta_0}(t(Z) > t_{obs}) \leq 1$, meaning we have at least one sample at each side of t_{obs}. The variance of an importance sample is $\sum w_i^2(f_i - E[f_i])^2$ (Owen, 2013, ch 9, pg 9), so a more equal distribution of weights would lead to better variances., a likely choice for θ_0 would often mean the selection criterion is not very strong, so that the samplers for θ_0 would be relatively efficient.

For choosing θ_0 under a truncation at c, one candidate choice is to use the unbiased estimator if there were no selection $\theta_0 = t_{obs}$. We require the empirical $0.2 < \hat{p}_{\theta_0} < 0.8$. If this value fails, we run a linear search for θ_0 values between t_{obs} and 0 until we find a successful value. Due to selection against small values of Z, t_{obs} is an upward biased estimator of θ; when variances are small, the sampler might need a better starting point for θ_0, so multiple starting points can be explored.

Samplers

Samplers for the truncated multivariate normal distribution are an area of active research, with new software available constantly. Examples available in the R environment include rejection samplers, Gibbs samplers (Geweke, 1991) and Hamiltonian (Pakman and Paninski, 2014) samplers. We use a Gibbs sampler that with an exponential approximation for extreme values.

2 Proof of Theorem 2

Theorem 2

Let $Z \sim MVN(\Theta, \Sigma)$, where $\Theta_{a:b} = \bar{\theta} \cdot \mathbf{s}$ for a pre-specified profile vector $\mathbf{s}$ and an unknown mean parameter $\bar{\theta}$. A confidence interval for $\bar{\theta}$ is estimated if $A_{a:b}$ occurs.

Define

$$l(Z) = \sup \left\{ \theta_l \,:\, t(Z) > \sup_{\theta < \theta_l} \{q_{1-\alpha}(\theta)\} \right\}, \quad u(Z) = \inf \left\{ \theta_u \,:\, t(Z) < \inf_{\theta > \theta_u} \{q_{\alpha}(\theta)\} \right\}, \quad (4)$$

For a Monte Carlo sample $\mathbf{z}_1, ..., \mathbf{z}_N$, let $l^{(N)}(Z) = l(\mathbf{z}_1, ..., \mathbf{z}_N)$ and $u^{(N)}(Z) = u(\mathbf{z}_1, ..., \mathbf{z}_N)$ be consistent estimators of $l(Z)$ and $u(Z)$.

Then the selective coverage of the interval $I^{(N)}(Z) = (l(Z)^{(N)}, u(Z)^{(N)})$ converges to at least $1 - 2 \cdot \alpha$

$$\lim_{N \to \infty} P\left(\bar{\theta} \in I^{(N)}(Z) \,|\, A_{a:b}\right) \geq 1 - 2 \cdot \alpha$$

Proof

For any θ,

$$lim_{N \to \infty} P(\theta \in [l(Z)^{(N)}, u(Z)^{(N)}] \,|\, A_{a:b}) = P(\theta \in [l(Z), u(Z)] \,|\, A_{a:b}),$$

because the density of $\{t(Z)|\,|\,A_{a:b}\}$ is bounded and $l(Z)^{(N)} \to l(Z),\ u(Z)^{(N)} \to u(Z)$.

Next, note that based on the definition of $u(Z)$, $q_\alpha(\theta) > t(Z)$ for any $\theta > u(Z)$. In particular, for the true parameter $\bar{\theta}$, the event $\bar{\theta} > u(Z)$ means that $q_\alpha(\bar{\theta}) > t(Z)$. Since $q_\alpha(\bar{\theta})$ is the α quantile of selective distribution $t(Z)\,|\,A_{a:b}$,

$$P\left(t(Z) < q_\alpha(\bar{\theta})\,|\,A_{a:b}\right) \leq \alpha,$$

so

$$P(\bar{\theta} > u(Z)\,|\,A_{a:b}) < \alpha.$$

A similar argument can be used to bound the probability of the event $\bar{\theta} < l(Z)$.

3 Proofs for the lemmas in Sections 3 and 4

Lemma 1

Let $Z \sim MVN(\Theta, \Sigma)$ and $Z' \in R^{a-b+3} \sim MVN(\Theta', \Sigma')$, with $\Theta' = \Theta_{a-1:b+1}$ and $\Sigma' = \Sigma_{a-1:b+1}$. Then

$$\{\,Z|A_{(a,b,+)}\}_{a-1:b+1} \stackrel{d}{=} \{Z'|A_{(2,a-b+2,+)}\}.$$

Proof

We need to show that $\{\,Z|A_{(a,b,+)}\}$ restricted to the coordinates $a-1, ..., b+1$ is a truncated multinormal vector with the shifted parameters Θ', Σ' and truncations according to $A_{(2,a-b+2,+)}$. Letting $\phi_D(\mathbf{z}, \Theta, \Sigma)$ be the D-dimensional MVN density, rewrite the conditional density $\{\,Z|A_{(a,b,+)}\}$ as

$$f_{Z|A}(\mathbf{z}) = \frac{\phi_D(\mathbf{z}, \Theta, \Sigma)}{\int_{A_{(a,b,+)}} \phi_D(\mathbf{u}, \Theta, \Sigma) d\mathbf{u}} 1(\mathbf{z} \in A_{(a,b,+)}).$$

The density of interest is therefore:

$$\begin{aligned}
f_{\{Z|A\}_{a-1:b+1}}(\mathbf{z}_{a-1:b+1}) = & \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \left[\frac{\phi_D(\mathbf{z}, \Theta, \Sigma)}{\int_{A_{(a,b,+)}} \phi_D(\mathbf{u}, \Theta, \Sigma) d\mathbf{u}} 1(\mathbf{z} \in A_{(a,b,+)}) \right] d\mathbf{z}_1 ... d\mathbf{z}_{a-2} d\mathbf{z}_{b+2} ... d\mathbf{z}_D \\
= & \frac{1(\mathbf{z}_{a-1:b+1} \in A_{(2,b-a+2,+)})}{\int_{A_{(a,b,+)}} \phi_D(\mathbf{u}, \Theta, \Sigma) d\mathbf{u}} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \left[\phi_D(\mathbf{z}, \Theta, \Sigma) \right] d\mathbf{z}_1 ... d\mathbf{z}_{a-2} d\mathbf{z}_{b+2} ... d\mathbf{z}_D \\
= & \frac{1(\mathbf{z}_{a-1:b+1} \in A_{(2,b-a+2,+)})}{\int_{A_{(a,b,+)}} \phi_D(\mathbf{u}, \Theta, \Sigma) d\mathbf{u}} \phi_{b-a+3}(\mathbf{z}_{a-1:b+1}, \Theta_{a-1:b+1}, \Sigma_{a-1:b+1}) \\
= & \frac{1(\mathbf{z}_{a-1:b+1} \in A_{(2,b-a+2,+)})}{\int_{A_{(a,b,+)}} \phi_D(\mathbf{u}, \Theta, \Sigma) d\mathbf{u}} \phi_{b-a+3}(\mathbf{z}_{a-1:b+1}, \Theta', \Sigma')
\end{aligned}$$

Furthermore, note that

$$\begin{aligned}
& \int_{A_{(a,b,+)}} \phi_D(\mathbf{u}, \Theta, \Sigma) d\mathbf{u} \\
& = \int_{-\infty}^{c} \int_{c}^{\infty} ... \int_{-\infty}^{c} \left[\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \frac{\phi_D(\mathbf{z}, \Theta, \Sigma)}{d} \mathbf{z}_1 ... d\mathbf{z}_{a-2} d\mathbf{z}_{b+2} ... d\mathbf{z}_D \right] d\mathbf{z}_{a-1} d\mathbf{z}_a ... d\mathbf{z}_{b+1} \\
& = \int_{-\infty}^{c} \int_{c}^{\infty} ... \int_{-\infty}^{c} \phi_{b-a+3}(\mathbf{z}_{a-1:b+1}, \Theta', \Sigma') d\mathbf{z}_{a-1} d\mathbf{z}_a ... d\mathbf{z}_{b+1} \\
& = \int_{A_{(2,b-a+2,+)}} \phi_{b-a+3}(\mathbf{z}_{a-1:b+1}, \Theta', \Sigma') d_{\mathbf{z}_{a-1:b+1}}
\end{aligned}$$

Lemma 2

Let $g_{\Theta(\theta;\mathbf{s})} := g_{\Theta(\theta)}$ denote the family densities for $t(Z)$ with a scale single parameter θ. Then

1. $g_{\Theta(\theta;\mathbf{s}_\Sigma)}$ is a monotone likelihood ratio family
2. $E[t(Z)]$ is an increasing function of θ
3. The confidence set for θ obtained by inverting two sided tests is an interval.

Proof

We recall the exponential family form of $f_{\Theta(\theta;\mathbf{s}_\Sigma)}$:

$$\begin{aligned} f_{Z \mid A[a,b];\Sigma}(\mathbf{z};\theta) = & \quad \exp\left\{\mathbf{z}\Sigma^{-1}\mu(\theta) + h(\mathbf{z}) - g(\theta)\right\} 1_A(\mathbf{z}), & (5) \\ = & \quad \exp\left\{\mathbf{z}\Sigma^{-1}(\theta\Sigma(1,...,1) + h(\mathbf{z}) - g(\theta)\right\} 1_A(\mathbf{z}) & (6) \\ = & \quad \exp\left\{\theta\mathbf{z}(1,...,1) + h(\mathbf{z}) - g(\theta)\right\} 1_A(\mathbf{z}) & (7) \end{aligned}$$

where $h(\mathbf{z}) = -\frac{1}{2}(\mathbf{z}\Sigma^{-1}\mathbf{z})$ does not depend on θ, and $g(\theta) = -\frac{1}{2}\mu(\theta)'\Sigma^{-1}\mu(\theta) - \int_A \phi(\mathbf{u}, \mu(\theta), \Sigma)d\mathbf{u}$ is the normalizing constant.

Next note that the density in $g_\Theta(x)$ for a particular value corresponds to integrating over the intersection of the hyper-plane $\sum Z = z$ with the conditional set A. Within this hyperplane, the value of the statistic is fixed $\mathbf{z}'(1,...,1) = x$ so

$$g_\theta(x) = \int_{A_x} \exp\{h(\mathbf{z})\} \exp\left\{\theta x - g(\theta)\right\} d\mathbf{z}.$$

Because the expression $\exp\left\{\theta x - g(\theta)\right\}$ is fixed within A_x, it can be moved outside the integral. Defining

$$H(x) = \int_{A_x} \exp\{h(\mathbf{z})\} d\mathbf{z},$$

we get the exponential family structure in

$$g_\theta(x) = H(x) \exp\left\{\theta x - g(\theta)\right\}.$$

Therefore, the likelihood ratio for a value $t(z) = x$ can be written as

$$\frac{g_{\theta_2(x)}}{g_{\theta_1(x)}} = \frac{H(x) \exp\left\{\theta_2 x - g(\theta_2)\right\}}{H(x) \exp\left\{\theta_1 x - g(\theta_1)\right\}} = \exp\{(\theta_2 - \theta_1)x - g(\theta_2) - g(\theta_1)\},$$

a strictly increasing function of x.

Lemma 3

We use results from Rinott and Scarsini (2006) for multivariate Normal distribution to identify the following conditions on Σ and $\mathbf{s}$:

1. Σ^{-1} is an M-matrix, meaning all off-diagonal elements are non-positive. (In particular, Σ must be non-negative).

2. $\mathbf{s}$ lies in the cone C_Σ of non-negative linear combinations of columns of Σ

$$C_\Sigma = \{\mu \in R^d : \mu'\Sigma^{-1} \geq \mathbf{0}\} = \{\mu \in \mathbf{R}^d : \text{ there exists } \mathbf{a} \geq \mathbf{0} \text{ s.t. } \mu = \mathbf{a}'\Sigma\}.$$

Note that for $\Sigma = I$, any non-negative profile would be in C_Σ (indeed, we see that all profiles produce a monotone curve for an iid covariance, see Figure 5 bottom-left in the main paper). The condition is sufficient, but not necessary; even if Σ is not an M-matrix, there might still be specific (profile,statistic) pairs for which monotonicity will hold, as discussed in Lemma 1. However, there will be no profile for which any statistic will be monotone.

Let $g_{\Theta(\theta);\mathbf{s}}$ be the family of densities of $t(Z)|A$ parametrized by θ, where $Z \sim N(\theta \cdot \mathbf{s}, \Sigma)$. If Σ^{-1} is an M-matrix, and the profile $\mathbf{s}$ can be written as a non-negative sum of columns of Σ, then $g_{\Theta(\theta;\mathbf{s})}$ is a monotone likelihood ratio family.

Proof

Consider random vectors Z and Z' where $Z \sim N(\theta \cdot \mathbf{s}, \Sigma)$ and $Z' \sim N(\theta' \cdot \mathbf{s}, \Sigma)$ for some $\theta' > \theta$. The lemma identifies a sufficient condition for $Z'|A$ being stochastically larger ($\geq_{st}$) than $Z|A$. Stochastic ordering implies that for any positive functional ϕ, $E[\phi(Z')|A] \geq E[\phi(Z)|A]$, and in particular for a positive statistic $t()$, $E[t(Z')|A] \geq E[t(Z)|A]$ and the quantile functions of $t(Z)|A$ are similarly ordered.

The key to the proof is moving from an ordering of Z and Z' into an ordering of the conditional vectors $[Z|A]$ and $[Z'|A]$, (interpreted as $[Z|Z \in A]$ and $[Z'|Z' \in A]$). For rectangular sets A, the stronger notion of ordering *total positivity* is maintained through the conditioning. We review here the main results of Rinott and Scarsini (2006); proofs and an extended discussion of the cited properties are found there.

- For multivariate densities f, g, the relation $f \preceq_{TP} g$ (total positivity) implies that $h(\mathbf{x}) := g(\mathbf{x})/f(\mathbf{x})$ is increasing in any coordinate-wise increase in $\mathbf{x}$. [Lemma 2.2, a]

- If $X \preceq_{TP} Y$, then X is also stochastically greater than Y, implying $E[\phi(X)] \leq E[\phi(Y)]$for any nondecreasing function ϕ [Proposition 2.4]. If $X \preceq_{TP} X$ then X is said to be *multivariate total positive of order 2* or MTP_2.

- For a rectangular set A, $X \preceq_{TP} Y$ implies $[X|X \in A] \preceq_{TP} [Y|Y \in A]$ [Theorem 2.5, a special condition of Remark 2.6 (i)] Note that conditioning by thresholding individual coordinates would always result in a rectangular set A.

- A multivariate normal Z with an invertible covariance matrix Σ is MTP2 if and only if Σ^{-1} is an M-matrix. [2.15]

- If for some $\mu \in \mathbf{R}^d$ we have $Z \preceq_{TP} Z + \mu$, then Z is MTP_2. [Thm 3.2]

- For Z that is MTP_2, we have $Z \preceq_{TP} Z + \mu$ *iff* $\mu \in C_\Sigma$ [Thm 3.2]

Therefore, under the assumptions regarding Σ, Z is $MTP-2$. Call $\mu = Z' - Z$, then $\mu = (\theta' - \theta) \cdot \mathbf{s}$ so $\mu \in C_\Sigma$ iff $\mathbf{s} \in C_\Sigma$. Therefore, the conditions suffice for $Z \preceq_{TP} Z'$. This further implies $[Z|A] \preceq_{TP} [Z'|A]$ and $E[\phi(Z)|A] \leq E[\phi(Z')|A]$.

4 Simulation and Data Experiments

Details of the simulation and data experiments are described below.

4.1 Simulation Experiment 1

In the first set of experiments, we tested coverage probability of $1 - 2\alpha = 0.9$ confidence intervals by repeatedly sampling the same set of variables, keeping only those data-vectors for which all locations passed the selection threshold. We sample data vectors of length $D = 5$, selecting for the *positive* region $r = (2, 4, +)$ spanning the vector. Samples were generated from the multivariate normal with mean region-heights of $\bar{\theta} = 0, 0.02, ..., 0.4$. For the true mean vector we used $\Theta = (\theta_1, \bar{\theta}, \bar{\theta}, \bar{\theta}, \theta_5)$ with $\theta_1 = \theta_5 = 0, \bar{\theta}/2$. Number of samples per group was $n_1 = n_2 = 4, 8, 16$; increased sample size reduces the variance of Z and, for unknown covariance, increases the accuracy of $\hat{\Sigma}$. The data was generated using two covariance matrices for the samples: a correlated C_{cor} and an uncorrelated C_{iid},

$$C_{cor} = \begin{pmatrix} 0.04 & 0.02 & 0.006 & 0 & 0 \\ 0.02 & 0.04 & 0.016 & 0 & 0 \\ 0.006 & 0.016 & 0.03 & 0 & 0 \\ 0 & 0 & 0 & 0.04 & 0.01 \\ 0 & 0 & 0 & 0.01 & 0.03 \end{pmatrix}, \quad C_{iid} = \begin{pmatrix} 0.04 & 0 & 0 & 0 & 0 \\ 0 & 0.04 & 0 & 0 & 0 \\ 0 & 0 & 0.04 & 0 & 0 \\ 0 & 0 & 0 & 0.04 & 0 \\ 0 & 0 & 0 & 0 & 0.04 \end{pmatrix}.$$

In both matrices, $\sigma^2 = 0.04$ was chosen to reflect the average observed within-group variance in the DNA-methylation data.

The following results are based on a tilting algorithm using $N = 12000$ samples from the reference distribution. The confidence interval estimation was repeated 1000 for each value each combination of parameters. We repeated each experiment three times, using (a)

the true covariance, (b) the sample covariance $\hat{\Sigma}$, and (c) the inflated covariance $\hat{\Sigma}_\lambda$ with $\lambda = 0.15$. We used the profile $\mathbf{s}_u$.

4.2 Simulation Experiment 2

In the second set of experiments, we verified coverage of the conditional estimator for selected regions within a continuous non-stationary process. Data was generated in the following way:

1. The (untransformed) mean process μ_A for group A was generated by sampling D=50 data points from an iid $N(0,1)$ process, and convolving with the 1-dimensional kernel $K_\mu = (0.1, 0.2, 0.4, 0.2, 0.1)$. The mean difference process μ_δ was generated by sampling iid $N(0, \sigma_\delta^2)$ and convolving with K_1. The mean for group B $\mu_B = \mu_A + \mu_\delta$.

2. For each sample we added noise $\varepsilon^i = (\varepsilon_1^i, ..., \varepsilon_{50}^i)$ by concatenating noise from two correlation regimes:

   - For $(\varepsilon_1^i, ..., \varepsilon_{D/2}^i)$, iid $N(0,1)$ samples were smoothed by convolving each vector with $K_\varepsilon = (0.05, 0.1, 0.15, 0.4, 0.15, 0.1, 0.05)$. This resulted in correlated noise with $\mathrm{E}[\varepsilon_j^i] = 0, \mathrm{Var}[\varepsilon_j^i] = \sigma_\varepsilon^2 = \|K_\varepsilon\|^2 = 0.23$.
   - For $(\varepsilon_{D/2+1}^i, ..., \varepsilon_D^i)$, iid noise was sampled from $N(0, \sigma_\varepsilon^2)$ with no smoothing.

3. Noise was added to each sample so that $Y^i = \mu_A + 1(i \in B) \cdot \mu_\delta + \varepsilon^i$.

4. For the transformed data, each sample Y^i was transformed coordinate-wise with the logistic function $\mathrm{logistic}(y) = \frac{\exp(y)}{1+\exp(y)}$. For the transformed data, a population of 10000 samples was generated, and the mean vector and covariance matrix for each group were estimated empirically from the samples.

5. Subsamples $n = 40, 20, 10, 5$ were taken from each group.

Results for each condition was based on 10 different populations, each sampled 100 times.

4.3 Simulation Experiment 3

In the third set, we wanted to see how power increases as we sample regions of increasing in length. We sampled vectors of length $D = 5, 7, 9, 11, 13, 15$, and selected for the region $r = (2, D-1, +)$ We used a flat mean $\Theta = (0, \bar{\theta}, ..., \bar{\theta}, 0)$ with mean effects size $\bar{\theta}$ ranging between 0 and 0.4. Covariances were block diagonal, with block sizes of 2 or 3 (as needed to get size D). Variances (of individual samples) were set to $\sigma^2 = 0.04$. Within the block correlation was $\rho = 0.2$, and outside the block correlation was $\rho = 0$. Each parameter combination was sampled 1000 times.

We ran our method with known covariance and a uniform profile. We compared this to the pivot method described in Lee et al. (2016), as implemented in the *selectiveInference* R package on CRAN (methods *mypoly.pval.lee* and *mypoly.int.lee*). We used $\alpha = 0.9$ for the intervals, and a one-sided $\alpha/2$ test to determine rejection.

4.4 Data Experiment

On each dataset, we detected ROIs using a fixed threshold ($c = 0.1$) to produce a list of candidate regions. For detection, we used the *bumphunter (v1.10.0)* package (Jaffe et al., 2012). For each detected ROI, we produced a selective p-value and formed a 90% selective confidence interval for mean between-group difference. The sample estimator of Σ was used for inference. Regions whose intervals overlapped 0 were pruned and intervals readjusted using (a) BH procedure to control FCR as discussed in Section 5 (main paper) or (b) Bonferroni procedure to control family-wise probability of non-coverage. The samples were

not smoothed or preprocessed. We did not allow regions to include sites separated by more than 5000 bps. The analysis was implemented in R. The TMN distribution was sampled using restrictMVN R package, and accelerated by tilting. For comparison, we produced p-values for the same set of detected ROIs using the family-wise error correction using permutations (Jaffe et al., 2012) as implemented in *bumphunter*. All candidate regions were ranked by area, and compared to the strongest regions found in a null distribution that assumes random assignment to groups (*fweArea*). The FWE corrected p-value was set to the proportion of permuted datasets in which a superior region was found.

5 Data information

	Colon Samples		Lung Samples
1	5775041068_R01C02	1	6929671086_R04C02
2	5775041068_R04C01	2	6285625064_R03C02
3	5775041068_R06C01	3	6285625087_R05C02
4	5775041068_R02C02	4	6285625087_R02C02
5	5775041068_R03C02	5	6285625087_R06C02
6	5775041068_R01C01	6	6285625099_R05C02
7	5775041068_R02C01	7	6285625099_R02C02
8	5775041068_R03C01	8	6285625090_R06C01
9	5775041068_R05C01	9	6285625087_R04C02
10	5775041088_R04C01	10	6929671122_R06C01
11	5775041084_R05C02	11	6285625090_R03C02
12	5775041065_R02C02	12	6285625090_R02C01
13	5775041088_R02C02	13	6264488083_R05C01
14	5775041065_R04C02	14	6285625090_R01C01
15	5775041007_R01C02	15	6285625099_R04C01
16	5775041084_R06C02	16	6285625064_R03C01
17	5775041084_R03C02	17	6285625099_R01C02
		18	6285625095_R01C01
		19	6285625090_R04C02

References

Fithian, W., Sun, D., and Taylor, J. (2014), "Optimal Inference After Model Selection," *arXiv preprint arXiv:1407.2812*, -.

Geweke, J. (1991), "Efficient simulation from the multivariate normal and student-t distri-

butions subject to linear constraints and the evaluation of constraint probabilities," in *Computing Science and Statistics*, Interface Foundation, pp. 571–578.

Jaffe, A. E., Murakami, P., Lee, H., Leek, J. T., Fallin, M. D., Feinberg, A. P., and Irizarry, R. A. (2012), "Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies," *International Journal of Epidemiology*, 41, 200–209.

Lee, J. D., Sun, D. L., Sun, Y., Taylor, J. E., et al. (2016), "Exact post-selection inference, with application to the lasso," *The Annals of Statistics*, 44, 907–927.

Owen, A. B. (2013), *Monte Carlo theory, methods and examples.*

Pakman, A. and Paninski, L. (2014), "Exact hamiltonian monte carlo for truncated multivariate gaussians," *Journal of Computational and Graphical Statistics*, 23, 518–542.

Rinott, Y. and Scarsini, M. (2006), "Total positivity order and the normal distribution," *Journal of Multivariate Analysis*, 97, 1251–1261.

Siegmund, D. (1976), "Importance sampling in the Monte Carlo study of sequential tests," *The Annals of Statistics*, 673–684.