

Selective inference is easier with p-values

Anav Sood
Stanford University

October 10, 2024

Abstract

Coming Soon

1 Introduction

Selective inference ... Modern selective methods however, which often condition on the selection event, can be difficult to derive, hard to implement, and often exhibit behaviors that is hard for non-experts to understand. In this work, we introduce selectively dominant p-values, a class of p-values that allow easy inference after selection. We show that these p-values have deep ties to many existing selective methods for doing inference on winners, and that rewriting these methods in terms of p-values leads easier derivations and deeper understandings of these methods. Our work also suggests new methods for more powerful global null testing, meta-analysis after selection, and inference on p-hacked p-values.

1.1 Our Contributions

Our work has a number of novel contributions.

Selectively dominant p-values:

Inference on the winner:

Hybrid inference:

Rank verification in exponential families:

More powerful null testing and correcting for publication bias:

1.2 Related Work

Related work (Andrews and Fithian on selection). (Adapt for increasing p-values). Other works on selecting the winner (Tijana, Benjamini)

Sequential selective hypothesis testing.

1.3 Notation

Stochastic dominance. when we say uniform we mean $\text{Unif}([0, 1])$. v_{-i} for a vector

Ties broken randomly. Mention something about dropping dependence on X .

Set minuses

2 Selectively Dominant p-Values

In this section we define selectively dominant p-values, a class of p-values that enable us to easily do inference after selection. We give a precise characterization of when p-values are selectively dominant and multiple examples illustrating that the most commonly used p-values are selectively dominant.

2.1 Selective Dominance

In classical statistics, a p-value is a random variable p that stochastically dominates the uniform distribution $U \sim \text{Unif}([0, 1])$ under the null, i.e., $p \succeq_{H_0} U$. When working with p-values, we maintain Type I error control if we reject the null H_0 when p is small:

$$P_{H_0}(\text{reject null}) = P_{H_0}(p \leq \alpha) \leq P(U \leq \alpha) = \alpha.$$

In essence, stochastic dominance allows us to use the uniform distribution as a reference distribution. We control the probability of p being small under the null by comparing it to the probability of a uniform being small.

In the problems we consider, we use p to test the null H_0 only after it has been “selected”. In full generality, we consider a p-value p that for testing the null H_0 that is conditionally valid given some random vector Z , i.e., $p \mid Z \succeq_{H_0} U$:

$$P_{H_0}(p \leq \alpha \mid Z) \leq P(U \leq \alpha) = \alpha. \quad (1)$$

Additionally, we consider a binary selection random variable $S \in \{0, 1\}$ that takes value one when p is selected. The relationship between p , Z , and S is governed by a **selection function**,

$$s(x, z) = P(S = 1 \mid p = x, Z = z),$$

that tells us the probability of selection given particular observed values of p and Z . We decide the selection procedure in our problems, so $s(x, z)$ is always known.

Because we only perform inference after selection, our goal is to design a procedure that controls Type I error conditional on selection:

$$P_{H_0}(\text{reject } H_0 \mid S = 1) \leq \alpha \quad (2)$$

As illustrated by our next example, the classical approach of rejecting when $p \leq \alpha$ does not maintain selective Type I error control as in (5).

Example 1 (Failure of classical inference). *Consider a p-value p that is uniform on $[0, 1]$ under the null H_0 . If we use the selection function $s(x) = I_{p \leq 1/2}$, i.e., we select p when it is at most $1/2$, then $p \mid S = 1 \sim \text{Unif}([0, 1/2])$. Our classical procedure will clearly fail to control Type I error conditional on selection:*

$$P_{H_0}(\text{reject } H_0 \mid S = 1) = P_{H_0}(p \leq \alpha \mid S = 1) = \min\{2\alpha, 1\} > \alpha.$$

Essentially, after selection, the uniform distribution no longer suffices as a reference distribution. Naturally, we may instead try and use the distribution of a uniform after it has been selected by the same selection function. Formally, suppose that U has uniform distribution conditional on Z , i.e., $U \mid Z \sim \text{Unif}([0, 1])$, and let $S \in \{0, 1\}$ be a binary selection random variable whose joint distribution with U and Z is governed by the same selection function

$$P(S = 1 \mid U = x, Z = z) = s(x, z).$$

Then, we could use the conditional distribution $U \mid Z, S = 1$ of U given selection as our reference distribution. This approach is valid exactly when our p-value is selectively dominant, as we define below.

Definition 1 (Selective dominance). *Considering a p-value p for the null H_0 that is valid given Z as in (1), we say that p is **selectively dominant given** Z if it has a conditional probability density function (PDF) under the null H_0 given Z and it satisfies*

$$p \mid Z, S = 1 \succeq_{H_0} U \mid Z, S = 1 \quad (3)$$

for every selection function $s(x, z)$ under which p and U both have a positive probability of being selected.

As we will soon see, the majority of p-values that practitioners use are selectively dominant as described in (3). In Definition 1, we restrict to p-values with conditional PDFs under the null because it makes our theory and methods simpler to state. Because we can always make a p-value both have a conditional PDF and be more powerful via randomization, this restriction is never a practical issue. Also, we only consider selection functions $s(x, z)$ under which p and U both have a positive probability of being selected so that the conditioning event $S = 1$ is never probability zero.

To perform valid post-selection inference using a selectively dominant p-value, we can transform it so that it remains a p-value after selection. As Theorem 1 explains, we can “undo” the effects of selection by applying the conditional cumulative distribution function (CDF) $F_{U|Z, S=1}(\cdot)$ of U given selection to p . Depending on the selection function $s(x, z)$, this CDF is often computable in closed form.

Theorem 1. *Let $F_{U|Z, S=1}(u)$ denote the CDF of U conditional on selection. Then, under the null, the transformed p-value*

$$p_{adj} = F_{U|Z, S=1}(p) = \frac{\int_0^p s(x, Z) dx}{\int_0^1 s(x, Z) dx} \quad (4)$$

stochastically dominates the uniform distribution conditional on Z and selection:

$$P_{H_0}(p_{adj} \leq \alpha \mid Z, S = 1) \leq \alpha. \quad (5)$$

Essentially, Theorem 1 tells us that if we want selective Type I error control as in (5), then we should reject H_0 when p is less than the α quantile of $U \mid Z, S = 1$ rather than the α quantile of U .

The below example illustrates how to apply Theorem 1 for a simple selection function. It illustrates how our selective dominance machinery allows us to easily test data dependent hypotheses, and it will come in handy when we discuss inference on winners.

Example 2 (Inference on the winning p-value). *Suppose we have n independent and selectively dominant p-values p_i for the null hypotheses $H_{0,i}$, and we choose to test only the j th null $H_{0,j}$ after observing that p_j is the smallest of the p_i . So long as p_j has a positive probability of being the smallest, then we can apply our framework with $p = p_j$, $Z = p_{-j}$, and the selection function $s(x, z) = I_{x < \min_k z_k}$. It is straightforward to compute that the adjusted p-value p_{adj} from (4) is $p_j / \min_{i \neq j} p_i$, so Theorem 1 tells us that rejecting when $p_j \leq \alpha \min_{i \neq j} p_i$ controls selective Type I error:*

$$P_{H_{0,j}}(p_j \leq \alpha \min_{i \neq j} p_i \mid p_{-j}, S = 1) \leq \alpha. \quad (6)$$

If we let W be the index of the smallest p-value, it is now easy to see that rejecting the data-dependent “winning” null $H_{0,W}$ when $p_{(1)} \leq \alpha p_{(2)}$ is a procedure that controls Type I error both conditionally on W and marginally. Consider only the set of indices $j \in \mathcal{J}$ for which p_j has a positive probability of being the smallest. Conditional error control is immediate: If $H_{0,j}$ is not true, then trivially $P(\text{falsely reject } H_{0,W} \mid W = j) = 0 \leq \alpha$. For the case that $H_{0,j}$ is true, the event $W = j$ is the same event as selecting p_j for inference in (6), so

$$\begin{aligned} P(\text{falsely reject } H_{0,W} \mid W = j) &= P(p_{(1)} \leq \alpha p_{(2)} \mid W = j) \\ &= P(p_j \leq \alpha \min_{i \neq j} p_i \mid W = j) \\ &\leq \alpha. \end{aligned}$$

Marginal error control follows from the law of total probability.

$$\begin{aligned} P(\text{falsely reject } H_{0,W}) &= \sum_{j \in \mathcal{J}} P(\text{falsely reject } H_{0,j} \mid W = j) P(W = j) \\ &\leq \alpha \sum_{j \in \mathcal{J}} P(W = j) \\ &\leq \alpha. \end{aligned}$$

If the nulls are all true and the p_i are exactly uniform, then the inequalities become equalities and our error control is tight.

2.2 Characterizing Selectively Dominant p-Values and Examples

Theorem 2 tells us that p-values are selectively dominant precisely when their conditional PDF is non-decreasing under the null.

Theorem 2 (Selective dominance and increasing density). *If the conditional PDF of the p-value p given Z is always non-decreasing under the null, then it is selectively dominant given Z as described in Definition 1. Conversely, if ever under the null, the conditional PDF of p given Z is everywhere continuous and not non-decreasing for a set of Z that have positive probability, then p is not selectively dominant given Z .*

In what follows, we give a number of examples of selectively dominant p-values. Our examples include all the common hypothesis testing scenarios that practitioners encounter in real life. We recommend that the unfamiliar reader review uniformly most powerful (UMP) and uniformly most powerful unbiased (UMPU) testing [Lehmann et al., 1986, Chapter 3 and Chapter 4] prior to proceeding.

Example 3 (Two-sided testing in parametric families). *Consider observing data from a parametric family P_θ and testing the null $H_0 : \theta = \theta_0$. Because the null is a point null, most p-values we construct will have an exact $\text{Unif}([0, 1])$ distribution under the null and are therefore trivially selectively dominant.*

Example 4 (One-sided testing in monotone likelihood ratio families). *Consider observing one-dimensional data from a parametric family $X \sim P_\theta$ that has density $p_\theta(x)$ with respect to some carrier measure μ . We say that P_θ has a monotone likelihood ratio (MLR) in the real valued function $T(x)$ if, for any $\theta < \theta'$, the ratio $p_{\theta'}(x)/p_\theta(x)$ is a finite non-decreasing function of $T(x)$. In this case, the UMP test for the null $H_0 : \theta \leq \theta_0$ rejects when $T(X)$ is large. The associated randomized p-value for this test is selectively dominant. We provide the explicit form of this p-value in Section 3.1.3 and a proof of this claim in ??.*

Example 5 (Testing in exponential families). *Suppose we observe data $X \in \mathbb{R}^m$ from an exponential family P_θ parameterized by $\theta \in \mathbb{R}^n$ i.e., under P_θ the data X has density*

$$g_\theta(x) = \exp(\theta_1 T_1(X) + \dots + \theta_n T_n(X) - \psi(\theta))g(x)$$

with respect to some carrier measure μ . In both the case of testing the two-sided null $H_0 : \theta_1 \neq \theta_{0,i}$ or one-sided null $H_0 : \theta_1 \leq \theta_{0,i}$, the UMPU test conditions on the nuisance statistics $T_{-i}(X)$. The p-value associated with the UMPU test for $H_0 : \theta_1 \neq \theta_{0,i}$ has an exact $\text{Unif}([0, 1])$ distribution conditional on $T_{-i}(X)$, so it is trivially selectively dominant given $Z = T_{-i}(X)$. For testing $H_0 : \theta_1 \leq \theta_{0,i}$, we are in the setting of an MLR family once we condition on $T_{-i}(X)$, so Example 4 implies that the p-value associated with the UMPU test is also selectively dominant given $Z = T_{-i}(X)$.

Example 6 (Permutation testing). *In a permutation test we observe data $X \in \mathcal{X}$ and compute a test statistic $T(X)$ that, under the null H_0 , has a distribution that is invariant under a finite group of transformations $G : \mathcal{X} \rightarrow \mathcal{X}$. That is, $T(X) \stackrel{d}{=}_{H_0} T(g(X))$ for all $g \in G$. To run the test, we consider a collection of group elements g_1, g_2, \dots, g_w where $g_1 = \text{id}$ is fixed to be the identity transformation and g_2, \dots, g_w are either a random sample from G with replacement or a random sample from $G \setminus \{\text{id}\}$ without replacement. The test then rejects when $T(X)$ is large compared to the $T(g_j(X))$. Specifically, the randomized permutation test from Proposition 3 of Hemerik and Goeman [2018] uses the p-value*

$$p = \frac{\#\{1 \leq j \leq w : T(g_j(X)) > T(X)\}}{w} + U_{aux} \frac{\#\{1 \leq j \leq w : T(g_j(X)) = T(X)\}}{w},$$

where $U_{aux} \sim \text{Unif}([0, 1])$ adds auxiliary randomness that is independent of X (not to be confused with the reference distribution U). This p-value always has an exact $\text{Unif}([0, 1])$ distribution under H_0 and is therefore trivially selectively dominant.

3 Inference on Winners

In this section we provide both new interpretations and generalizations of methods that perform inference on winners. Our main insight is that, once written in terms of p-values, inference on winners methods become easier to derive, interpret, implement, and adapt to new settings.

Our treatment covers three methods for doing inference on winners: classical inference, conditional inference from [Fithian et al. \[2017\]](#), and hybrid inference from [Andrews et al. \[2023\]](#). For each method, we first discuss how the method goes about constructing a lower confidence bound (LCB) for the winning mean in the simple setting of independent Gaussian data. This discussion mirrors how the methods are typically presented in the existing literature. Then we write the LCB in terms of p-values and discuss the benefits of doing so.

3.1 Classical Inference

In this subsection we discuss how to perform classical inference on the winner. The material is review, but it lays the foundation of what is to come.

3.1.1 The Classical LCB

In the case of independent Gaussian data $X \sim N(\mu, I_n)$ with winner $W = \operatorname{argmax}_{i \in [n]} X_i$, the classical approach to getting an LCB for the winning mean μ_W is to perform inference simultaneously on all the μ_i . Defining

$$\alpha_n = 1 - (1 - \alpha)^{1/n}, \quad (7)$$

the lower bounds $\mu_i > X_i - z_{1-\alpha_n}$ ¹ hold simultaneously with high probability:

$$P(\mu_i > X_i - z_{1-\alpha_n} \text{ for all } i \in [n]) = P(\max_{i \in [n]} X_i - \mu_i < z_{1-\alpha_n}) = P(X_1 - \mu_1 < z_{1-\alpha_n})^n = 1 - \alpha.$$

Therefore $X_W - z_{1-\alpha_n}$ is a strict lower bound for μ_W with probability at least $1 - \alpha$.

Because this classical LCB arises from performing simultaneous inference on all the parameters, it necessarily suffers a penalty as the number of parameters increases. Figure 1 depicts how the quantile $z_{1-\alpha_n}$ grows with the number of parameters n . As a consequence, the classical lower bound for the winning mean μ_W becomes progressively farther away from the observed value X_W as the problem dimension increases.

3.1.2 The Classical LCB with p-Values

To gauge the quality of our classical LCB, we can check when it verifies the existence of a mean larger than $\mu_0 \in \mathbb{R}$. In the language of hypothesis testing, we are considering the nulls $H_{0,i}^{\mu_0} : \mu_i \leq \mu_0$ and rejecting the data dependent winning null $H_{0,W}^{\mu_0}$ when our LCB is at least μ_0 .

We can check when the classical LCB allows us to reject $H_{0,W}^{\mu_0}$ by examining the p-values $p^{\mu_0}(X_i) = 1 - \Phi(X_i - \mu_0)$, which are valid p-values under the nulls $H_{0,i}^{\mu_0}$. It is not hard to verify that the classical LCB rejects $H_{0,W}^{\mu_0}$ when the winning (smallest) p-value $p_{(1)}^{\mu_0}$ is at most α_n . The classical LCB, which is composed of all the μ_0 for which we fail to reject $H_{0,W}^{\mu_0}$, can thus be written as

$$C_{\text{class}}(X) = \{\mu_0 : p_{(1)}^{\mu_0} > \alpha_n\}. \quad (8)$$

When working with independent p-values, rejecting nulls whose corresponding p-value is less than α_n is a well-known family wise error rate (FWER) controlling procedure called Sidak's procedure. We recall it in Theorem 3 [[Šidák, 1967](#)]. Essentially, the classical LCB (8) for μ_W results from inverting Sidak's test for $H_{0,W}$.

Theorem 3 (Sidak's procedure). *Suppose that p_i are independent p-values for the nulls $H_{0,i}$. If we reject $H_{0,i}$ when $p_i \leq \alpha_n$ then the FWER is controlled at level α . As a consequence, if W is the index of the smallest p-value (with ties broken arbitrarily), then rejecting the winning null $H_{0,W}$ or the global null $H_0 : \cap_{i=1}^n H_{0,i}$ when $p_{(1)} \leq \alpha_n$ controls Type I error at level α .*

One benefit reframing things in a multiple hypothesis testing framework is that, if we are not explicitly interested in making a confidence region for μ_W and care just about making discoveries, we can make more discoveries via the closure principle [[Marcus et al., 1976](#)]. To close Sidak's procedure, we imagine applying

¹When α is small or n is large, $\alpha_n \approx \alpha/n$, so this approach is essentially equivalent to doing inference on each mean individually with a Bonferroni correction.

Sidak's global null test for every subseted global null $H_{0,\mathcal{I}}^{\mu_0} = \cap_{i \in \mathcal{I}} H_{0,i}^{\mu_0}$ where $\mathcal{I} \subseteq [p]$. We then reject the individual null $H_{0,i}^{\mu_0}$ if and only if Sidak's procedure rejects $H_{0,\mathcal{I}}^{\mu_0}$ for every \mathcal{I} containing i . If the subseted global null tests are level α , then our individual rejections will satisfy level α FWER control.

Closing a global null test always allows us to make more discoveries, but doing so is usually computationally intractable. The closure of Sidak's procedure, however, has a tractable characterization called Holm's procedure. We give it in Corollary 1.

Corollary 1 (Holm's procedure). *Suppose that p_i are independent p-values for the nulls $H_{0,i}$. As shorthand, let $H_{0,(j)}$ denote the null corresponding to the j th smallest p-value, with ties broken arbitrarily. Rejecting the null hypothesis $H_{0,(k)}$ when $p_{(j)} \leq \alpha_{n-j+1}$ for every $j \leq k$ controls FWER error at level α .*

Holm's procedure is best understood as a sequential procedure. We reject $H_{0,(1)}$ if $p_{(1)} \leq \alpha_n$. Then, if we rejected $H_{0,(1)}$, we reject $H_{0,(2)}$ if $p_{(2)} \leq \alpha_{n-1}$, so on and so forth until we fail to reject.

3.1.3 Generalizing to MLR Families

A big advantage of writing everything in terms of p-values is that it easily allows us to generalize to any MLR family. Consider any family of distributions P_θ parameterized by $\theta \in \mathbb{R}$ with an MLR in $T(x)$, and suppose we observe n independent samples $X_i \sim P_{\theta_i}$ where the θ_i are unknown. This is a strict generalization of our independent Gaussian setting.

Denoting the left-continuous survival function of $T(X)$ and its righthand limit as

$$G^{\theta_0}(t) = P_{\theta_0}(T(X) \geq t) \quad G^{\theta_0,+}(t) = \lim_{u \downarrow t} G^{\theta_0}(u)$$

the p-value corresponding to the UMP test for $H_{0,i}^{\theta_0} : \theta_i \leq \theta_0$ is

$$p^{\theta_0}(X_i) = G^{\theta_0,+}(T(X_i)) + U_{i,aux}(G^{\theta_0}(T(X_i)) - G^{\theta_0,+}(T(X_i))). \quad (9)$$

In (9), $U_{i,aux}$ are $\text{Unif}([0, 1])$ random variables that are independent from each other and the X_i . We give a detailed explanation regarding why rejecting according to these p-values is UMP in ??.

Corollary 2 tells us that, if we swap out our Gaussian p-values for the UMP p-values in (9), (8) directly generalizes to give us simultaneously valid LCBs for the θ_i . Since the LCBs are simultaneously valid, we automatically get an LCB for the winning parameter(s) from them.

Corollary 2 (Classical inference for MLR families). *Let P_θ be a MLR family parameterized by $\theta \in \mathbb{R}$. If $X_i \sim P_{\theta_i}$ are n independent samples and $p_i^{\theta_0}$ are the UMP p-values (9) for testing $H_{0,i}^{\theta_0} : \theta_i \leq \theta_0$, then $\{\theta_0 : p_i^{\theta_0} > \alpha_n\}$ are a collection of LCBs for the θ_i that simultaneously hold with probability exactly $1 - \alpha$.*

3.2 Conditional Inference

In this subsection we provide the same treatment for conditional inference on the winner. The p-value viewpoint makes the existing method much easier to interpret and implement, and it also results in novel methods and generalizations.

3.2.1 The Conditional LCB

Again suppose we observe independent Gaussian data $X \sim N(\mu, I_n)$ and want to make a LCB for the mean μ_W of the winner $W = \text{argmax}_{i \in [n]} X_i$.

The conditional approach tries to avoid the classical approach's curse of dimensionality by providing inferences for only the winning mean. It does so by delivering an LCB for μ_W that is valid conditionally on W . To construct the conditional LCB, we follow the framework of Fithian et al. [2017]. Letting R be the index of the runner-up (second largest observation), we note that the deviation of X_W from μ_W has a truncated normal distribution once we condition on W and X_{-W} :

$$X_W - \mu_W \mid W, X_{-W} \sim TN(0, 1, X_R - \mu_W, \infty). \quad (10)$$

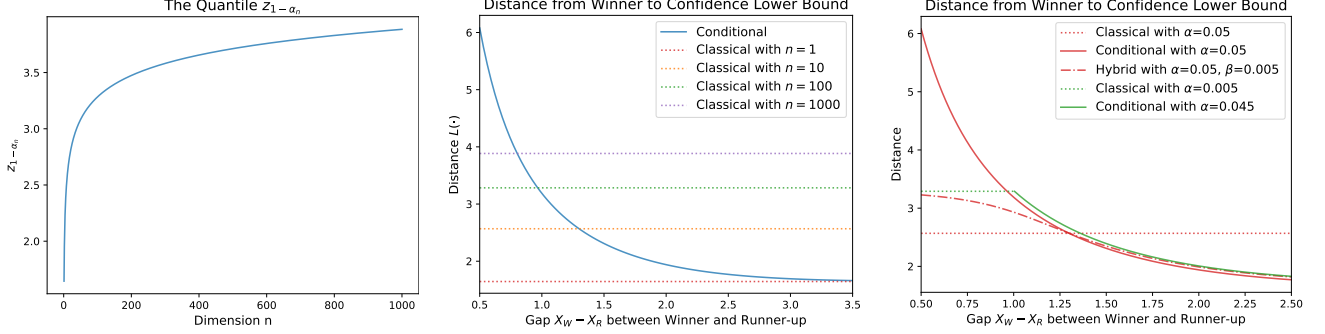


Figure 1: The first panel (left) shows the growth of the quantile $z_{1-\alpha_n}$ as a function of the dimension n . The second panel (middle) gives the distance between the level $\alpha = 0.05$ conditional LCB and the winner X_W (i.e., the function $L(\cdot)$ in (13)) as a function of the gap $X_W - X_R$ between winner and the runner-up. For dimensions $n = 1, 10, 100, 1000$, we also give the distance from X_W to the level $\alpha = 0.05$ classical LCB. The third panel (right) considers a $n = 10$ dimensional problem and shows this distance for the level $\alpha = 0.05$ hybrid, conditional, and classical LCBs in red. For hybrid we take $\beta = 0.005$. It also shows the distance to the larger of the level $\alpha = 0.045$ conditional LCB and level $\alpha = 0.005$ classical LCB in green (i.e., the union bound LCB).

Letting

$$q_{1-\alpha}(w, x_{-w}, \mu_w) = \text{Quantile}_\mu(1 - \alpha, X_W - \mu_W \mid W = w, X_{-w} = x_{-w}) \quad (11)$$

denote the $1 - \alpha$ quantile of this conditional distribution (10), it is straightforward to show that

$$C_{\text{cond}}(X) = \{\eta : \eta > X_W - q_{1-\alpha}(W, X_{-W}, \eta)\} \quad (12)$$

is a $1 - \alpha$ confidence region for μ_W that is valid conditional on W and X_{-W} :

$$P_\mu(\mu_W \in C_{\text{cond}}(X) \mid W, X_{-W}) = P_\mu(X_W - \mu_W < q_{1-\alpha}(W, X_{-W}, \mu_W) \mid W, X_W) = 1 - \alpha.$$

Since η appears on both sides of the membership condition in (12), it is not clear that $C_{\text{cond}}(X)$ actually provides a lower bound for μ_W . We will see, however, that we can rewrite $C_{\text{cond}}(X)$ as

$$C_{\text{cond}}(X) = \{\eta : \eta > X_W - L(X_W - X_R)\}. \quad (13)$$

where $L : \mathbb{R} \rightarrow \mathbb{R}$ is a complicated “length” function that determines the distance from the winner X_W to the lower confidence bound. We will de-mistify $L(\cdot)$ in the next sub-section. For now, we provide a plot of it in the middle panel of Figure 1.

As can be seen in Figure 1, the conditional LCB has very interesting behavior. In some situations, it indeed seems to avoid the curse of dimensionality. If the gap between X_W and X_R is large, then the conditional LCB for μ_W will be roughly $X_W - z_{1-\alpha}$, i.e, what we expect in a one-dimensional inference problem. But the cost of using the conditional approach can be tremendous. As the runner-up gets close to the winner, the conditional LCB goes quickly to $-\infty$ and can give much worse inferences than the classical approach. For example, if the runner-up is within just one standard deviation of the winner, we get inferences similar to those in the $n = 100$ dimensional classical problem.

Compared to the classical approach, which provides a lower bound that is a fixed distance from the winner, the conditional approach is much harder to interpret. Its quality depends on a somewhat opaque relationship between the runner-up and the winner. How separated does the winner need to be from the runner-up for the conditional approach to be useful? And how close can the winner and runner-up be before the classical approach is better? We shed light on these questions by rewriting the conditional procedure in terms of p-values.

3.2.2 The Conditional LCB with p-Values

As we did for the classical LCB, imagine rejecting the winning null $H_{0,W}^{\mu_0} : \mu_W \leq \mu_0$ when the conditional LCB is at least μ_0 . Again, we will check this condition by examining the p-values $p_i^{\mu_0} = 1 - \Phi(X_i - \mu_0)$. To be more concrete, we will assume $\alpha = 0.05$ from now onwards.

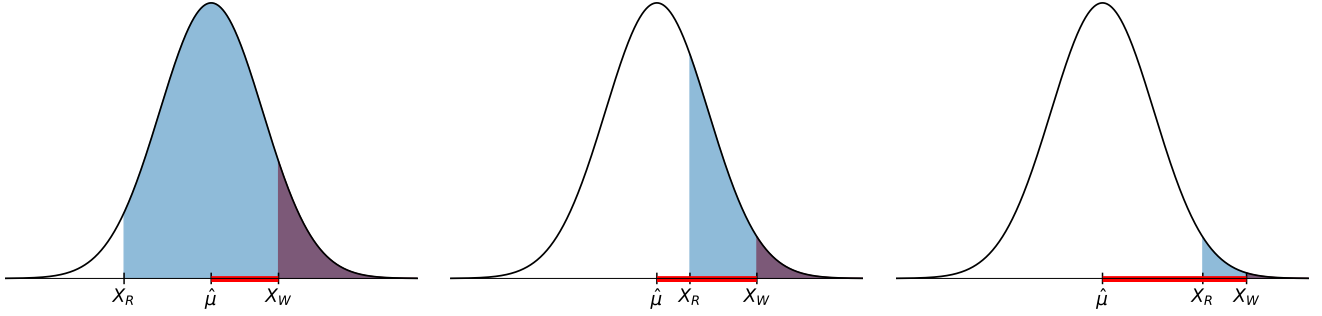


Figure 2: We plot the level $\alpha = 0.1$ conditional LCB $\hat{\mu}$ for different gaps between the winning value X_W and the runner up value X_R and highlight the distance between $\hat{\mu}$ and X_W in red. The LCB $\hat{\mu}$ is chosen exactly so that the tail probability $P(N(\hat{\mu}, 1) > X_R)$, shaded in blue, is $1/\alpha = 10$ times the tail probability $P(N(\hat{\mu}, 1) > X_W)$, shaded in red (the overlap appears purple). As X_W and X_R get closer, we need to take $\hat{\mu}$ further and further back for this condition to be satisfied.

In contrast to the classical LCB, the conditional LCB rejects $H_0^{\mu_0}$ when the winning (smallest) p-value $p_{(1)}^{\mu_0}$ is small relative to the runner-up (second smallest) p-value $p_{(2)}^{\mu_0}$. Particularly, it rejects when $p_{(1)}^{\mu_0} \leq \alpha p_{(2)}^{\mu_0}$. We can therefore write the conditional LCB (12), which is composed of all the μ_0 for which we fail to reject $H_{0,W}^{\mu_0}$, as

$$C_{\text{cond}}(X) = \{\mu_0 \in \mathbb{R} : p_{(1)}^{\mu_0} > \alpha p_{(2)}^{\mu_0}\}. \quad (14)$$

Derivations verifying these facts can be found in Appendix A.4.

From (14) we get a clear interpretation of the length function $L(\cdot)$ from (13). The conditional LCB stretches all the way back to the mean μ_0 under which it is $1/\alpha = 20$ times less likely to see something as extreme as the winner than something as extreme as the runner-up! It is not hard to show (see Appendix A.1) that the distance between X_W and the conditional LCB is determined solely by the gap between X_W and X_R . If the winner and runner-up are very close, we will need μ_0 to be very far back. Thanks to the decay of the Gaussian tail, however, we can always find such a mean if we go far back enough. Figure 2 provides an illustration.

As we have already seen in Example 2, when working with independent and selectively dominant p-values, rejecting the winning null when the smallest p-value is at most α times the second smallest p-value is an error controlling procedure. Equipped with our selective dominance machinery, the computations in Example 2 were quite straightforward. In place of the previous sub-section's computations, it would have been much easier to justify our conditional LCB as an inversion of Example 2's test for the winner. For completeness, we restate the result from Example 2 in Theorem 4.

Theorem 4 (Testing the winner). *Suppose that p_i are n independent and selectively dominant p-values under the nulls $H_{0,i}$, and let W be the index of the smallest p-value. Rejecting $H_{0,W}$ when $p_{(1)} \leq \alpha p_{(2)}$ controls Type I error at level α conditionally on W , and therefore also marginally. The same is thus true for rejecting the global null $\cap_{i=1}^n H_{0,i}$ when $p_{(1)} \leq \alpha p_{(2)}$.*

Once written in terms of p-values, its easy to see the merits and pitfalls of the conditional approach mathematically. If all the p-values except the smallest provide essentially no evidence against the null (i.e, the winner is well separated), then $p_{(2)} \approx 1$ and we reject when $p_{(1)} < \alpha$, the same p-value cutoff as a one-dimensional problem. On the other hand, if even just one other p-value provides a similar amount of evidence against the null as the smallest p-value, then $p_{(2)} \approx p_{(1)}$ and we will never reject because $p_{(1)} \not\leq \alpha p_{(1)} \approx \alpha p_{(2)}$. In general, the conditional approach only makes a discovery if the most extreme observation is $1/\alpha = 20$ more extreme under the null than the second most extreme observation. This can be quite a stringent requirement!

The p-value viewpoint also makes it clear that the conditional approach can only outperform the classical one when some null p-values are conservative. In our Gaussian problem, the p-value for $H_{0,i}^{\mu_0} : \mu_i \leq \mu_0$ is conservative when $\mu_i \ll \mu_0$. We give a heuristic argument here, and a more formal statement in Appendix A.2.

Suppose one of our p-values p_1 is a very strong signal (i.e., $\mu_1 \gg \mu_0$) but the remaining p-values p_2, \dots, p_n are non-conservative nulls (i.e., $\mu_2, \dots, \mu_n = \mu_0$). Our conditional procedure will reject when our smallest p-value, likely p_1 , is less than α times the minimum of p_2, \dots, p_n , which are $n - 1$ independent uniform random variables that have a minimum of $1/n$ on average. Thus, roughly speaking, the conditional approach also rejects when $p_{(1)}$ is less than $\alpha/n \approx \alpha_n$, which is no better than Sidak's.

Like was the case for our classical procedure, we can make more discoveries via the conditional approach by closing Theorem 4's global null test. Although the resulting procedure is a bit strange, it is interesting that Theorem 4's global null test admits a tractable closure.

Corollary 3 (Closed testing for winners). *Suppose that p_i are n independent and selectively dominant p-values under the nulls $H_{0,i}$. As shorthand, let $H_{0,(j)}$ denote the null corresponding to the j th smallest p-value and define $p_{(n+1)} = 1$. Rejecting the null hypotheses $H_{0,(k)}$ for which $p_{(j)} \leq \alpha p_{(j-1)}$ for every $j \leq k$ controls FWER error at level α .*

This closed procedure is also best understood sequentially. We reject $H_{0,(1)}$ if $p_{(1)} \leq \alpha p_{(2)}$. Then, if we rejected $H_{0,(1)}$, we reject $H_{0,(2)}$ if $p_{(2)} \leq \alpha p_{(3)}$, so on and so forth until we fail to reject.

3.2.3 Generalizing to MLR Families

Written in terms of p-values, the conditional LCB (14) also easily generalizes to any MLR family. Suppose P_θ is an MLR family and we observe n independent samples $X_i \sim P_{\theta_i}$ where the θ_i are unknown. Let $p_i^{\theta_0}$ be the UMP p-value for testing $H_{0,i} : \theta_i \leq \theta_0$ as in (9). Because the auxiliary randomness we use in (9) is shared across different θ_0 , the index of the smallest p-value will be the same for every θ_0 . Hence, we can unambiguously define the winning index $W = \operatorname{argmin}_{i \in [n]} p_i^{\theta_0}$ to be the index of the smallest p-value.

Corollary 4 tells us how to make an LCB for the winning parameter θ_W that is valid conditionally on W .

Corollary 4 (Conditional inference for MLR families). *Let P_θ be a MLR family parameterized by $\theta \in \mathbb{R}$. If $X_i \sim P_{\theta_i}$ are n independent samples and $p_i^{\theta_0}$ is the UMP p-value (9) for testing $H_{0,i} : \theta_i \leq \theta_0$, then*

$$\{\theta_0 : p_{(1)}^{\theta_0} > \alpha p_{(2)}^{\theta_0}\} \quad (15)$$

is an LCB for θ_W that holds conditionally on W (and therefore also marginally) with probability exactly $1 - \alpha$.

The proof of Corollary 4 is essentially that the proposed LCB results from inverting the test of the winning null $H_{0,W}^{\theta_0} : \theta_W \leq \theta_0$ from Theorem 4.

We point out that the distance from the winning value X_W to the conditional LCB (15) will depend on quickly the tail of P_θ decays. If the the tail decays faster, then the first θ_0 for which the largest observation is $1/\alpha = 20$ times as extreme as the second largest observation will be larger. Seeing as the Gaussian distribution, which has an exponentially decaying tail, still often results in very low lower bounds, we should expect that this distance can often be quite large.

3.3 Hybrid Inference

Lastly we provide the same analysis for hybrid inference, which tries to balance the benefits of the classical and conditional approaches.

3.3.1 The Hybrid LCB

Like before, we want to make a LCB for the mean μ_W of the winner $W = \operatorname{argmax}_{i \in [n]} X_i$ in the case of independent Gaussian data $X \sim N(\mu, I_n)$ with unknown mean $\mu \in \mathbb{R}^n$.

The core idea behind hybrid inference is giving a confidence region $C_{hyb}(X)$ that has a very high probability of containing μ_W on a "good" event G_μ . Oddly, this good event depends on the unknown parameter. For some $\beta < \alpha$, we need G_μ to happen with probability at least $1 - \beta$. Then, if we ensure that $C_{hyb}(X)$

has at least $(1 - \alpha)/(1 - \beta)$ coverage on the G_μ , it will achieve $1 - \alpha$ coverage overall:

$$\begin{aligned} P_\mu(\mu_W \in C_{hyb}(X)) &= P_\mu(G_\mu)P_\mu(\mu_W \in C_{hyb}(X)|G_\mu) + P_\mu(G_\mu^c)P_\mu(\mu_W \in C_{hyb}(X)|G_\mu^c) \\ &\geq (1 - \beta) \left(\frac{1 - \alpha}{1 - \beta} \right) \\ &= 1 - \alpha. \end{aligned}$$

Considering some $\beta < \alpha$ and defining $\beta_n = 1 - (1 - \beta)^{\frac{1}{n}}$ as in (7), our good event is that the confidence lower bounds $X_i - z_{1-\beta_n}$ for the means μ_i all simultaneously hold:

$$G_\mu = \{X_i < \mu_i + z_{1-\beta_n} \text{ for all } i \in [n]\}.$$

From our earlier reasoning, we know that this good event happens with probability exactly $1 - \beta$.

Now, we can make a confidence region that contains the mean with probability at least $(1 - \alpha)/(1 - \beta)$ on this good event. If we condition on G_μ along with W and X_{-W} , the deviation of X_W from μ_W has a truncated normal distribution like (10) that is further truncated from above:

$$X_W - \mu_W \mid W, X_{-W}, G_\mu \sim TN(0, 1, X_W - \mu_W, z_{1-\beta_n}). \quad (16)$$

On the good event G_μ , we always have that $X_R - \mu_W < X_W - \mu_W < z_{1-\beta_n}$, so the lower truncation is indeed below the upper one. Let

$$q_{\frac{1-\alpha}{1-\beta}}^h(w, x_{-w}, \mu_w) = \text{Quantile}_\mu \left(\frac{1-\alpha}{1-\beta}, X_W - \mu_W \mid W = w, X_{-w} = x_{-w}, G_\mu \right) \quad (17)$$

denote the $(1 - \alpha)/(1 - \beta)$ quantile of the conditional distribution (16). Per the prior discussion, the function (17) only makes sense if the largest value in x_{-w} at most $z_{1-\beta_n}$, and we will take the quantile (17) to be $-\infty$ if it is not. It is then straightforward to show that

$$C_{hyb}(X) = \{\eta : \eta > X_W - q_{\frac{1-\alpha}{1-\beta}}^h(W, X_{-W}, \eta)\} \quad (18)$$

contains μ_W with high probability conditional on W , X_{-W} , and the event G_μ :

$$P_\mu(\mu_W \in C_{hyb}(X) \mid W, X_{-W}, G_\mu) = P_\mu(X_W - \mu_W < q_{\frac{1-\alpha}{1-\beta}}^h(W, X_{-W}, \mu_W) \mid W, X_{-W}, G_\mu) = \frac{1 - \alpha}{1 - \beta}.$$

Based on our earlier discussions, this is sufficient to imply that $C_{hyb}(X)$ from (18) will contain μ_W with probability at least $1 - \alpha$.

As was the case for our conditional confidence region (12), the hybrid confidence region (18) is a little hard to interpret at first. We will get a clearer sense of its benefits when we write it in terms of p-values. As a teaser, the rightmost panel of Figure 1 plots distance between the winner X_W and the hybrid LCB for a $n = 10$ dimensional problem. We argue (using our upcoming p-value viewpoint) in Appendix A.3 that, once we fix α and β , this distance is a function of just the gap $X_W - X_R$ between the winner and runner up and the problem dimension n .

3.3.2 The Hybrid LCB with p-Values

As we did for the conditional LCB, imagine rejecting the winning null $H_{0,W}^{\mu_0} : \mu_W \leq \mu_0$ when the hybrid LCB is at least μ_0 . Again, we will check this condition by examining the p-values $p_i^{\mu_0} = 1 - \Phi(X_i - \mu_0)$.

It turns out that the hybrid LCB allows us to reject $H_{0,W}^{\mu_0}$ when the winning (smallest) p-value is smaller than a mixture of the runner-up (second smallest) p-value and the level β classical cutoff:

$$p_{(1)}^{\mu_0} \leq \frac{\alpha - \beta}{1 - \beta} p_{(2)}^{\mu_0} + \left(1 - \frac{\alpha - \beta}{1 - \beta} \right) \beta_n \quad (19)$$

Hence, we can write the hybrid LCB (18) as

$$C_{hyb}(X) = \left\{ \mu_0 \in \mathbb{R} : p_{(1)}^{\mu_0} > \frac{\alpha - \beta}{1 - \beta} p_{(2)}^{\mu_0} + \left(1 - \frac{\alpha - \beta}{1 - \beta} \right) \beta_n \right\}. \quad (20)$$

The derivations needed to show this can be found in Appendix A.5.

Theorem 5 tells us that, when our p-values are independent and selectively dominant, rejecting the most promising null according the hybrid cutoff (19) is a Type I error controlling procedure. The derivations required to show this are similar to those in Example 2, but a bit more involved, so we have moved the proof to the appendix.

Theorem 5 (Hybrid test for the winner). *Suppose that p_i are n independent and selectively dominant p-values under the nulls $H_{0,i}$, let W be the index of the smallest p-value, and fix some $\beta < \alpha$. Rejecting $H_{0,W}$ when*

$$p_{(1)} \leq \frac{\alpha - \beta}{1 - \beta} p_{(2)} + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_n \quad (21)$$

controls Type I error at level α . The same is thus true for rejecting the global null $\cap_{i=1}^n H_{0,i}$ whenever (21) holds.

Once written in terms of p-values, it is much easier to see how the hybrid approach balances the benefits of the classical and conditional approaches. When the other p-values provide essentially no evidence against the null (i.e., $p_{(2)} \approx 1$), hybrid rejects when $p_{(1)}$ is less than a cutoff that is bounded away from zero. Hence, when the winner is well separated, the hybrid approach avoids the curse of dimensionality, at least on par with the conditional procedure run at level $(1 - \alpha)/(1 - \beta)$. On the other hand, even if some other p-value provides as much evidence against the null as the smallest, hybrid still rejects whenever the level β classical LCB does. This is because when $p_{(1)} \leq \beta_n$, the hybrid cutoff is a mixture of two things that are larger than $p_{(1)}$, and we will reject. When we set $\beta = 0$ then the hybrid cutoff (21) becomes $\alpha p_{(2)}$ and we recover the conditional method, and if we set $\beta = \alpha$ it becomes α_n and we recover the classical method.

Interestingly, the hybrid global null test from Theorem 5 also admits a tractable closure that allows us to make more discoveries.

Corollary 5 (Closed hybrid testing for winners). *Suppose that p_i are n independent and selectively dominant p-values under the nulls $H_{0,i}$. As shorthand, let $H_{0,(j)}$ denote the null corresponding to the j th smallest p-value and define $p_{(n+1)} = 1$. Fixing some $\beta < \alpha$, rejecting the null hypotheses $H_{0,(k)}$ for which*

$$p_{(j)} \leq \frac{\alpha - \beta}{1 - \beta} p_{(j-1)} + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_{n-j+1}$$

for every $j \leq k$ controls FWER error at level α .

This closed procedure is also best understood sequentially. We reject $H_{0,(1)}$ if $p_{(1)} \leq \frac{\alpha - \beta}{1 - \beta} p_{(2)} + (1 - \frac{\alpha - \beta}{1 - \beta}) \beta_n$. Then, if we rejected $H_{0,(1)}$, we reject $H_{0,(2)}$ if $p_{(2)} \leq \frac{\alpha - \beta}{1 - \beta} p_{(3)} + (1 - \frac{\alpha - \beta}{1 - \beta}) \beta_{n-1}$, so on and so forth until we fail to reject.

3.3.3 Comparison to Union Bound

In contrast to hybrid inference, a naive union bound approach rejects the winning null whenever the level β classical approach rejects or the level $\alpha - \beta$ conditional approach rejects, i.e., it rejects whenever

$$p_{(1)} \leq \max\{(\alpha - \beta)p_{(2)}, \beta_n\}. \quad (22)$$

The union bound harshly switches between the classical and conditional approaches, whereas the hybrid approach smoothly interpolates between them. This is illustrated in the rightmost panel of Figure 1, which compares the LCBs resulting from the hybrid versus union bound approaches.

Written in terms of p-values, we easily see that the hybrid approach dominates the union bound approach. Both methods reject when $p_{(1)} \leq \beta_n$, and when $p_{(1)} > \beta_n$ the hybrid cutoff (21) is clearly strictly larger than the union bound cutoff (22), meaning hybrid will reject whenever the union bound does and more. There are two reasons hybrid inference out performs the union bound. First, union bounds are not tight and waste some of the Type I error budget. This is captured by the fact that the coefficient on $p_{(2)}$ is larger in the

hybrid cutoff (21) than the union bound cutoff (22). Second, conditioning on the “good event” in the hybrid approach causes an upper truncation in (16), which results in $\left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_n$ being added to the cutoff ².

Practically speaking, however, hybrid inference typically results in very little improvement over the union bound. This is already somewhat evident in Figure 1’s rightmost panel, where we see that the hybrid LCB and union bound LCB are very similar. Even when the hybrid cutoff is larger than that of the union bound, it is provably not much larger. We detail why in Appendix A.6, where we also run a number of simulations comparing the power of the hybrid and union bound approaches. We are hard pressed to find a setting where the hybrid approach results in a meaningful power gain. Hence, we suggest adopting the view that hybrid inference’s main contribution is squeezing out the little power that is left from the union bound approach. As it is not computationally more expensive and our p-value viewpoint makes it equally easy to implement, we feel it is still worth using.

3.3.4 Generalizing to MLR Families

Now written in terms of p-values, the hybrid LCB for the winning parameter (20) immediately generalizes to any MLR family. Like before, suppose P_θ is an MLR family and we observe n independent samples $X_i \sim P_{\theta_i}$ where the θ_i are unknown. Again, let $p_i^{\theta_0}$ be the UMP p-value for testing $H_{0,i} : \theta_i \leq \theta_0$ as in (9) and let $W = \operatorname{argmin}_{i \in [n]} p_i^{\theta_0}$ be the index of the smallest p-value.

Corollary 6 tells us how to make a hybrid LCB for the winning parameter θ_W . For the same reasons as in the Gaussian setting, the hybrid LCB continues to balance the benefits of the conditional and classical approaches in the more general MLR setting.

Corollary 6 (Hybrid inference for MLR families). *Let P_θ be a MLR family parameterized by $\theta \in \mathbb{R}$ and fix some $\beta < \alpha$. If $X_i \sim P_{\theta_i}$ are n independent samples and $p_i^{\theta_0}$ is the UMP p-value (9) for testing $H_{0,i}^{\theta_0} : \theta_i \leq \theta_0$, then*

$$\left\{ \theta_0 : p_{(1)}^{\theta_0} \leq \frac{\alpha - \beta}{1 - \beta} p_{(2)}^{\theta_0} + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_n \right\} \quad (23)$$

is an LCB for the winning parameter θ_W that holds with probability at least $1 - \alpha$.

Again, the proof of Corollary 6 is essentially that the proposed LCB is the inversion of the test of the winning null $H_{0,W}^{\theta_0} : \theta_W \leq \theta_0$ from Theorem 5.

4 Rank Verification in Exponential Families

In this section we consider the problem of verifying that the winning parameter is actually larger than the other parameters, i.e., rather than doing inference on the winning parameter, we do inference on the gap between the winning parameter and the remaining parameters. Mainly, we give a corrected accounting of the seminal work Hung and Fithian [2019]’s main results in the language of selected dominance. It illustrates how our selective dominance machinery allows us to correctly design fairly intricate statistical procedures.

4.1 Warm-up: Rank Verification and Type III Error Control

To motivate the rank verification problem and shed some light on how it relates to selective dominance, we consider a seemingly unrelated classical statistical question about Type III errors.

A friend of yours tests whether the unknown means of two univariate Gaussian samples, $X_1 \sim N(\mu_1, 1/\sqrt{2})$ and $X_2 \sim N(\mu_2, 1/\sqrt{2})$, are different. They end up rejecting the null hypothesis $H_0 : \mu_1 = \mu_2$ because the two-sided UMPU p-value $1 - \Phi(|X_1 - X_2|)$ is at most α . After rejecting, they note $X_1 > X_2$, and claim “not only am I confident that the two means are different, but they must be different because μ_1 is bigger than μ_2 ”. Your friend, however, only rejected the null that the means are equal. Can they make a claim about the direction of inequality? This is a question of Type III error, and we can use our selective dominance framework to show that your friend’s claims are actually statistically valid.

²the authors Andrews et al. [2023] mention these differences between the union bound and hybrid inference in a footnote, but they only point out that hybrid dominates the level β classical approach, which is weaker than our statement.

Based on their claim, it seems that what your friend really wants to do is test the one-sided null $H_{0,1} : \mu_1 \leq \mu_2$ whenever they observe that $X_1 > X_2$, and test the complementary one-sided null $H_{0,2} : \mu_2 \leq \mu_1$ whenever they observe that $X_2 < X_1$. To test the null $H_{0,j} : \mu_j \leq \mu_{-j}$ we normally use the UMP p-value $1 - \Phi(X_j - X_{-j})$. In your friend's case, however, they only select this p-value to use for inference when they observe that $X_j > X_{-j}$, or equivalently that $p_j < 1/2$. Example 7 illustrates how, using our selective dominance framework, your friend could develop a procedure that delivers valid inferences in spite of this selection.

Example 7 (Rank verification in a simple case). *Suppose that p is a selectively dominant p-value for testing the null H_0 , but we only choose to test H_0 when $p < 1/2$. Applying our framework with the p-value p and selection function $s(x) = I_{x < 1/2}$, Theorem 1 tells us that we control selective Type I error if we reject according to the adjusted p-value from (4) is $p_{adj} = 2p$:*

$$P_{H_0} \left(p_j \leq \frac{\alpha}{2} \mid S = 1 \right) \leq \alpha. \quad (24)$$

Now, consider data $X_1 \sim N(\mu_1, 1/\sqrt{2})$ and $X_2 \sim N(\mu_2, 1/\sqrt{2})$, the one-sided nulls $H_{0,j} : \mu_j \leq \mu_{-j}$, and their corresponding selectively dominant p-values $p_j = 1 - \Phi(X_j - X_{-j})$ (selective dominance follows from Example 4). Denoting the winner $W = \operatorname{argmax}_{j=1,2} X_j$, it is now clear rejecting the data-dependent null $H_{0,W} : \mu_W \leq \mu_{-W}$ when $p_W \leq \alpha/2$ maintains Type I error control both conditionally on W and marginally. If $H_{0,j}$ is not true, then trivially $P(\text{falsely reject } H_{0,W} \mid W = j) = 0 \leq \alpha$. For the case that $H_{0,j}$ is true, the event $W = j$ is identical to the event $p_j < 1/2$, and hence is the same event as selecting p_j for inference in (24). Therefore,

$$\begin{aligned} P(\text{falsely reject } H_{0,W} \mid W = j) &= P \left(p_W \leq \frac{\alpha}{2} \mid W = j \right) \\ &= P \left(p_j \leq \frac{\alpha}{2} \mid W = j \right) \\ &\leq \alpha, \end{aligned}$$

implying error control conditional on W . Marginal error control then follows from the law of total probability:

$$\begin{aligned} P(\text{falsely reject } H_{0,W}) &= \sum_{j=1,2} P(\text{falsely reject } H_{0,j} \mid W = j) P(W = j) \\ &\leq \alpha \sum_{j=1,2} P(W = j) \\ &= \alpha. \end{aligned}$$

If $\mu_1 = \mu_2$ then the inequalities become equalities and our error control is tight.

In (7), we provide a procedure that verifies the rank of the winning mean with Type I error control, i.e., it affirms not just that the means are different, but that the winning mean μ_W is larger than the other mean. The procedure, which rejects when the smaller of the two one-sided p-values $p_i = 1 - \Phi(X_i - X_{-i})$ is at most $\alpha/2$, is identical to the procedure that rejects when the two-sided p-value $1 - \Phi(|X_1 - X_2|)$ is at most α ! Hence, for reasons likely unbeknownst to them, your friend's original claim is indeed statistically valid.

In what follows, we generalize (7)'s rank verification procedure to the exponential family setting. Although this generalization appears quite a bit more complicated, the core idea remains the same.

4.2 Rank Verification in Exponential Families

In this sub-section we illustrate how to do rank verification when we observe data $X \in \mathbb{R}^m$ from a multiparameter exponential family P_θ with density

$$g_\theta(x) = \exp(\theta_1 T_1(x) + \cdots + \theta_n T_n(x) - \psi(\theta)) g(x) \quad (25)$$

We typically care about doing inference on some subset $\mathcal{S} \subseteq [p]$ of the θ_j (e.g., in the case of a multivariate Gaussian with unknown covariance $\sigma^2 I_n$, we usually are only interested in θ_j that determine the means and

not the θ_j that determines the variance). Since $E_\theta[T] = \theta$, it makes sense to let the winner W be the index in \mathcal{S} corresponding to the largest sufficient statistic T_i , with ties broken randomly.

Considering the nulls $H_{0,ij}^\delta : \theta_i - \theta_j \leq \delta$, we want to reject the data-dependent global null $\cap_{j \in \mathcal{S}-W} H_{0,Wj}^\delta$ and affirm that θ_W is more than δ larger than any other parameter in \mathcal{S} . Then we can make an LCB for the difference $\theta_W - \max_{j \in \mathcal{S}-W} \theta_j$ between the winning and next largest parameter in \mathcal{S} by inverting this test.

Considering some $i \neq j$, we start by constructing the UMPU p-value $p_{ij}^\delta(X)$ for testing $H_{0,ij}^\delta : \theta_i - \theta_j \leq \delta$. Defining the transformed sufficient statistics $\tilde{T} \in \mathbb{R}^n$ by

$$\tilde{T}_i = \frac{T_i - T_j}{2}, \quad \tilde{T}_j = \frac{T_i + T_j}{2}, \quad \tilde{T}_\ell = T_\ell \text{ for } \ell \neq i, j \quad (26)$$

we can reparametrize the density (27) as

$$g_\theta(x) = \exp \left((\theta_i - \theta_j) \tilde{T}_i + (\theta_i + \theta_j) \tilde{T}_j + \sum_{\ell \neq i, j} \tilde{T}_\ell \theta_\ell \right) g(x). \quad (27)$$

From this reparameterization, it is clear that the conditional left-continuous survival function of \tilde{T}_i and its righthand limit

$$G_{ij}^\delta(\tilde{t}_i | \tilde{t}_{-i}) = P_{\theta_i - \theta_j = \delta}(\tilde{T}_i \geq \tilde{t}_i | \tilde{T}_{-i} = \tilde{t}_{-i}) \quad G_{ij}^{\delta,+}(\tilde{t}_i | \tilde{t}_{-i}) = \lim_{u \downarrow \tilde{t}_i} G_{ij}^\delta(u | \tilde{t}_{-i}), \quad (28)$$

have no dependence on θ , and the UMPU p-value p_{ij}^δ for testing $H_{0,ij}^\delta : \theta_i - \theta_j \leq \delta$ is given by

$$p_{ij}^\delta(X) = G_{ij}^{\delta,+}(\tilde{T}_i | \tilde{T}_{-i}) + U_{ij,aux}(G_{ij}^{\theta_0}(\tilde{T}_i | \tilde{T}_{-i}) - G_{ij}^{\theta_0,+}(\tilde{T}_i | \tilde{T}_{-i})), \quad (29)$$

where $U_{ij,aux}$ are $\text{Unif}([0, 1])$ random variables that are independent from each other and X (see Chapter Four of [Lehmann et al. \[1986\]](#) and ?? for a justification).

Crucially, we can tell if $i \in \mathcal{S}$ is a winner by examining p_{ij}^δ . It is straightforward to confirm that $i \in \mathcal{S}$ is the sole winner exactly when $\tilde{T}_i > \max_{k \in \mathcal{S}-i} \tilde{T}_k - \tilde{T}_j$. Equivalently, this happens when p_{ij}^δ is strictly smaller than

$$q_{ij}^{\delta,+}(\tilde{T}_{-i}) = G_{ij}^{\delta,+}(\max_{k \in \mathcal{S}-i} \tilde{T}_k - \tilde{T}_j | \tilde{T}_{-i}). \quad (30)$$

Likewise, one can confirm that $i \in \mathcal{S}$ is one of multiple winners in \mathcal{S} exactly when $\tilde{T}_i = \max_{k \in \mathcal{S}-i} \tilde{T}_k - \tilde{T}_j$, or equivalently when p_{ij}^δ is at least $q_{ij}^{\delta,+}$ but at most

$$q_{ij}^\delta(\tilde{T}_{-i}) = G_{ij}^\delta(\max_{k \in \mathcal{S}-i} \tilde{T}_k - \tilde{T}_j | \tilde{T}_{-i}), \quad (31)$$

Moreover, in the case that there are multiple winners, the number of winners is also a deterministic function of \tilde{T}_{-i} :

$$N_i(\tilde{T}_{-i}) = 1 + \{k \in \mathcal{S} - i : \tilde{T}_k = \max_{\ell \in \mathcal{S}-i} \tilde{T}_\ell\}. \quad (32)$$

Note that $N_i(\tilde{T}_{-i})$, which is always at least two, is not the same as the number of winners, which can be one. Rather, it is the number of winners there will be if \tilde{T}_i is a winner and at least one other \tilde{T}_j is as well.

Leveraging these facts, [Example 8](#) shows how we can use our selective dominance framework to come up with valid tests for both the winning null H_{Wj}^δ and the winning global null $\cap_{j \in \mathcal{S}-W} H_{0,Wj}^\delta$.

Example 8. Suppose p is a p-value for the null H_0 that is selectively dominant given Z . If we select p to use for inference according to the selection function

$$s(x, z) = \begin{cases} 1 & \text{if } x < q^+(z), \\ \frac{1}{N(z)} & \text{if } x \in [q^+(z), q(z)], \\ 0 & \text{otherwise,} \end{cases}$$

where $N(z) > 1$ and $0 \leq q^+(z) \leq q(z) \leq 1$ are known functions of z , then the adjusted p -value from (4) turns out to be (see Appendix A.7 for computations)

$$p_{adj} = f(p, q^+(Z), q(Z), N(Z)) \quad f(a, b, c, d) = \frac{a - (1 - \frac{1}{d})(a - b)_+}{b + \frac{1}{d}(c - b)}. \quad (33)$$

Therefore, Theorem 1 tells us that rejecting when (33) is at most α is a selective Type I error controlling procedure:

$$P_{H_0} (f(p, q^+(Z), q(Z), N(Z)) \leq \alpha \mid Z, S = 1) \leq \alpha. \quad (34)$$

Now, suppose we observe X drawn from the exponential family (27) and let W be the index $i \in \mathcal{S}$ of the largest sufficient statistic (with ties broken randomly). For $i \neq j$, Example 5 tells us that the UMPU p -value $p = p_{ij}^\delta$ from (29) for the null $H_{0,ij}^\delta : \theta_i - \theta_j \leq \delta$ is selectively dominant given the transformed nuisance statistics $Z = \tilde{T}_{-i}$ from (26). Taking $q^+(Z) = q_{ij}^{\delta,+}(\tilde{T}_{-i})$, $q(Z) = q_{ij}^\delta(\tilde{T}_{-i})$, $N(Z) = N_i(\tilde{T}_{-i})$ and f from (30), (31), (32), and (33), it is now easy to show that rejecting the data-dependent null $H_{0,Wj}^\delta$ when $f(p_{Wj}^\delta, q_{Wj}^{\delta,+}, q_{Wj}^\delta, N_{Wj}) \leq \alpha$ controls Type I error both conditional on W and marginally. Again it suffices to restrict our attention to indices $i \in \mathcal{I}$, $i \neq j$ that have a positive probability of being the winner (if $i = j$, then we know H_{ii}^0 is true and we simply do not reject). If $H_{0,ij}^\delta$ is not true, then trivially $P(\text{falsely reject } H_{0,Wj}^\delta \mid W = i) = 0 \leq \alpha$. For the case that $H_{0,Wj}^\delta$ is true (and $i \neq j$), the event $W = i$ is the same event as selecting $p_{0,ij}^\delta$ for inference in (33), so

$$\begin{aligned} P(\text{falsely reject } H_{0,Wj}^\delta \mid W = i) &= P(f(p_{Wj}^\delta, q_{Wj}^{\delta,+}, q_{Wj}^\delta, N_{Wj}) \leq \alpha \mid W = i) \\ &= P(f(p_{ij}^\delta, q_{ij}^{\delta,+}, q_{ij}^\delta, N_i) \leq \alpha \mid W = i) \\ &\leq \alpha. \end{aligned}$$

Marginal error control follows from the usual law of total probability argument.

We can reject the data-dependent global null when $\cap_{j \in \mathcal{S}-W} H_{0,Wj}^\delta$ when we reject all of the individual nulls $H_{0,Wj}^\delta$ for $j \in \mathcal{S} - W$. It is straightforward to see that this will control Type I error both conditionally on W and marginally: If there is an $i \in \mathcal{I}$ for which $\cap_{j \in \mathcal{S}-i} H_{0,ij}^\delta$ is false, then trivially $P(\text{falsely reject } \cap_{j \in \mathcal{S}-W} H_{0,Wj}^\delta \mid W = i) = 0 \leq \alpha$ for this i . Otherwise, there exists a $k \in \mathcal{S} - i$ for which $\theta_i \leq \theta_k$, and

$$P(\text{falsely reject } \cap_{j \in \mathcal{S}-W} H_{0,Wj}^\delta \mid W = i) \leq P(\text{falsely reject } H_{0,Wk}^\delta \mid W = i) \leq \alpha.$$

Again, marginal error control follows from the usual law of total probability argument.

By inverting Example 8's test for the data-dependent global null $\cap_{j \in \mathcal{S}-W} H_{0,Wj}^\delta$, we can get our desired LCB for $\theta_W - \max_{j \in \mathcal{S}-W} \theta_j$. We give a precise statement in Theorem 6 below.

Theorem 6. Let X be drawn from the exponential family (27) and let W be the index $i \in \mathcal{S}$ corresponding to the largest sufficient statistic T_i (with ties broken randomly). If p_{ij}^δ is the UMPU p -value (29) for testing $H_{0,ij}^\delta : \theta_i - \theta_j \leq \delta$ and $q_{ij}^{\delta,+}$, q_{ij}^δ , and N_i are as in (30), (31), and (39), then

$$\left\{ \delta : \max_{j \in \mathcal{S}-W} \frac{p_{Wj}^\delta - \left(1 - \frac{1}{N_i}\right) (p_{Wj}^\delta - q_{Wj}^{\delta,+})_+}{q_{Wj}^{\delta,+} + \frac{1}{N_i} (q_{Wj}^\delta - q_{Wj}^{\delta,+})} > \alpha \right\} \quad (35)$$

is a LCB for θ_W that holds with probability at least $1 - \alpha$ conditionally on W (and therefore also marginally).

Our derivation of the test for $H_{0,Wj}$ differs from that given in Hung and Fithian [2019] in two important ways. First, the derivation in Hung and Fithian [2019] only implies Type I error control at the boundary of the one-sided null, although they assume it to be true for the entire composite null. Due to selection, establishing validity for entire composite null is non-trivial. Thankfully, it is immediately implied by our selective dominance machinery and Example 5. Second, Hung and Fithian [2019] incorrectly use the same adjusted p -value when there are and are not ties amongst the T_j . If there cannot be ties amongst the T_j ,

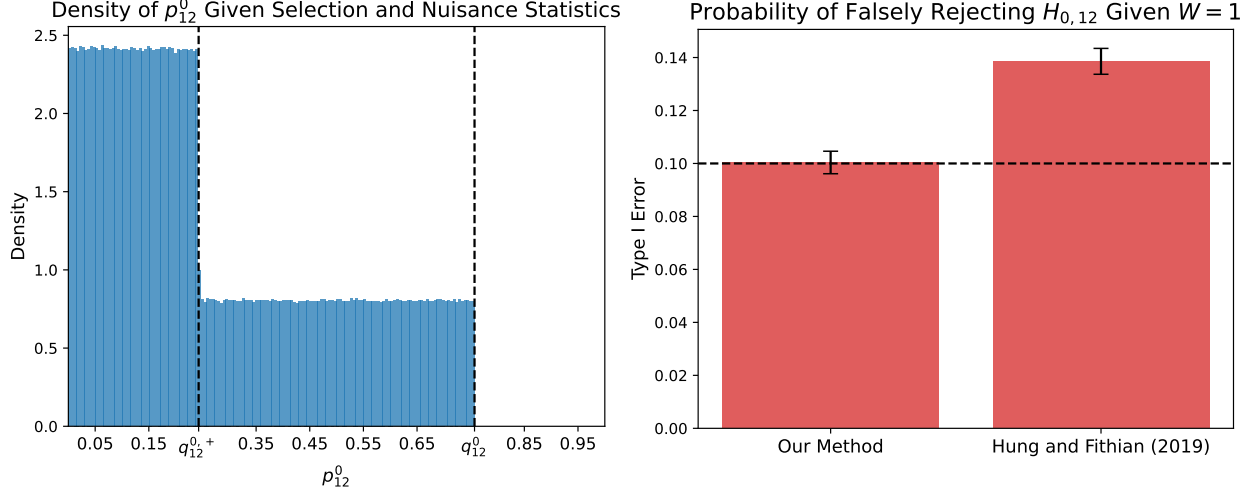


Figure 3: Considering three independent binomial random variables $X_i \sim \text{Bin}(n, s_i)$ with random variables with $n = 4$ and $s_i = 1/2$, the first panel (left) depicts $N = 10^6$ draws from the conditional distribution of the p-value p_{12}^0 (used for testing $H_{0,12}^0 : s_1 \leq s_2$) given selection $W = 1$ and the nuisance statistics $(X_1 + X_2)/2 = 4$, $X = 3$. When $p_{12}^0 < q_{12}^{0,+}$, then X_1 is the sole winner and the p-value is selected for inference with probability one, but when $p_{12}^0 \in [q_{12}^{0,+}, q_{12}^0]$ there is a three-way tie and it is selected for inference with probability $1/3$. Hence, the p-value's conditional distribution is not uniform on $[0, q_{12}^0]$ as [Hung and Fithian \[2019\]](#) implicitly assume. The next panel (right) displays the consequence. Conditional on $W = 1$, [Hung and Fithian \[2019\]](#) do not maintain Type I error control when testing H_{W2}^0 at level $\alpha = 0.1$ (denoted by horizontal dashed line), whereas our method does. Error bars denote a 99% confidence interval.

then $q_{ij}^\delta = q_{ij}^{\delta,+}$ and the adjusted p-value (33) we use to test $H_{0,ij}^\delta$ conditional on $W = i$ simplifies to $p_{ij}^\delta/q_{ij}^\delta$. If we use $p_{ij}^\delta/q_{ij}^\delta$ as our adjusted p-value when ties are possible, we will not achieve conditional error control (as is claimed in [Fithian et al. \[2017\]](#)). In the case that $X \in \mathbb{R}^3$ is three independent binomials, the left panel of Figure 3 depicts the conditional distribution of p_{12}^0 given $W = 1$ and a specific setting of the nuisance statistics Z . It makes it clear that rejecting when $p_{ij}^\delta/q_{ij}^\delta < \alpha$ does not maintain error control, as affirmed in Figure 3's right panel.

In closing, we mention that the final result (Theorem 6) is perhaps not as surprising as [Hung and Fithian \[2019\]](#) originally suggest. In [Hung and Fithian \[2019\]](#), there are further computations showing that if the exponential family density (27) is such that $T(X) = X$, the function $g(x)$ is Schur concave (see [Fithian et al., 2017](#), Definition 2), and there are no ties, then (1) the maximum in (35) is achieved by the runner up R (the index $j \in S$ with second largest sufficient statistic T_j) and (2) q_{WR}^0 is equal to $1/2$. Hence, in this restricted setting, we can claim that the winning parameter is indeed the largest whenever $p_{WR}^0 \leq \alpha/2$, i.e., whenever the two-sided test comparing the winner and runner-up rejects. Since this test has non-vanishing level even as n grows, the authors claim that it seems to surprisingly circumvent a multiple comparisons issue. However, Example 7 already illustrates why verifying rank by running a two-sided test comparing the winner and loser gives Type I error control in the $n = 2$ dimensional case. It should be obvious then that running a two-sided test to compare the winner to the best of $n - 1$ losers in the $n > 2$ dimensional case is a more conservative procedure. In other words, the multiplicity correction does not appear in the level of the test, but it is baked in by the fact that we compare the winner to the largest of $n - 1$ different losers, which scales with n .

4.3 Inference on Winners in Exponential Families

We can use the same tools from the previous sub-section to do inference on the winner in multiparameter exponential families. Again consider data X drawn from the exponential family (27) and let the winner W be the index $i \in S \subseteq [p]$ corresponding to the largest sufficient statistic T_i , with ties broken randomly.

The p-value

$$p_i^{\theta_0}(T(X)) = G_i^{\theta_0,+}(T_i|T_{-i}) + U_{i,aux}(G_i^{\theta_0}(T_i|T_{-i}) - G_i^{\theta_0,+}(T_i|T_{-i})), \quad (36)$$

corresponds to the UMP test for $H_{0,i}^{\theta_0} : \theta_i \leq \theta_0$, where, like before,

$$G_i^{\theta_0}(t_i|t_{-i}) = P_{\theta_0}(T_i \geq t_i | T_{-i} = t_{-i}) \quad G_i^{\theta_0,+}(t) = \lim_{u \downarrow t} (T_i \geq t | T_{-i} = t_{-i}),$$

denote the conditional left-continuous survival function of T_i and its right-hand limit, and $U_{i,aux}$ are $\text{Unif}([0, 1])$ random variables that are independent from each other and X .

To come up with a test for the data dependent null $H_{0,W}^{\theta_0}$ (which we will then invert to get an LCB for the winning parameter), we need to apply a similar adjustment to our p-value as in Example 8. Rather than re-work through essentially identical arguments, we simply define the analogous quantities to those from the previous sub-section,

$$q_i^{\theta_0,+}(T_{-i}) = G_i^{\theta_0,+}(\max_{k \in S_{-i}} T_k | T_{-i}), \quad (37)$$

$$q_i^{\theta_0}(T_{-i}) = G_i^{\theta_0}(\max_{k \in S_{-i}} T_k | T_{-i}), \quad (38)$$

$$N_i(T_{-i}) = 1 + |\{k \neq i : T_k = \max_{\ell \in S_{-i}} T_\ell\}|, \quad (39)$$

and state the final result in Corollary 7. In ?? we use a modification of the arguments in Fithian et al. [2017] to show that our test $H_{0,W}^{\theta_0}$ is selectively UMPU, meaning it is UMPU amongst all tests that are valid conditional on W . Hence, as stated in Corollary 7, its inversion is selectively uniformly most accurate unbiased (UMAU) in the analogous sense.

Corollary 7 (Conditional inference for multiparameter exponential families). *Let X be drawn from the exponential family (27) and let W be the index $i \in S$ corresponding to the largest sufficient statistic T_i (with ties broken randomly). If $p_i^{\theta_0}$ is the UMP p-value (36) for testing $H_{0,i}^{\theta_0} : \theta_i \leq \theta_0$ and $q_i^{\theta_0,+}$, $q_i^{\theta_0}$, and N_i are as in (37), (38), and (39), then*

$$\left\{ \theta_0 : \frac{p_W^{\theta_0} - \left(1 - \frac{1}{N_i}\right)(p_W^{\theta_0} - q_W^{\theta_0,+})}{q_W^{\theta_0,+} + \frac{1}{N_i}(q_W^{\theta_0} - q_W^{\theta_0,+})} > \alpha \right\} \quad (40)$$

is an LCB for θ_W that holds with probability exactly $1 - \alpha$ conditional on W (and therefore also marginally), and it is selectively UMAU conditional on W .

Briefly, we point out that selecting the largest sufficient statistic does not always correspond to selecting the “most promising” effect. Even though $T_i > T_j$, it might be the case that $p_j^{\theta_0}$ is smaller than $p_i^{\theta_0}$. Unlike in Section 3, we cannot simply have the winner correspond to the smallest p-value for two reasons. First, in this more general setting, the index resulting in the smallest p-value changes depending on θ_0 (see Appendix A.8 for an example). Hence, it is not well-defined to determine the winner by looking at the smallest p-value (and we thus have to deal with ties). Second, even if we settled on a θ_0 to prioritize, because the p-values have intricate correlation structure, the selection event $p_i^{\theta_0} < \max_{j \neq i} p_j^{\theta_0}$ corresponds to a very complicated selection function that we would have to try and characterize on a case-by-case basis. Indeed, this discussion applies to the rank verification problem for exponential families we discussed previously as well.

5 New Selective Methods

In this section we illustrate how our selective dominance framework allows us to easily design new methods for selective inference, global null testing, and dealing with publication bias.

Rather than just selecting just one p-value as in Section 2, some of this section’s methods select multiple p-values to use for inference. As such, we need to slightly generalize our framework from Section 2. The generalization is intuitive, and for sake of brevity we have deferred a formal account of it to Appendix B. The proofs of validity for this section’s methods, which are in the appendix, often rely on Appendix B.

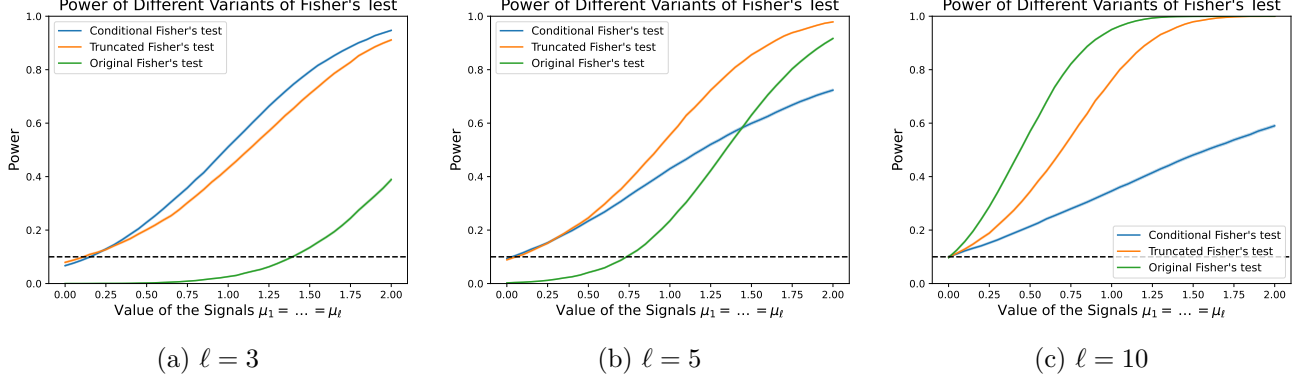


Figure 4: For $\ell = 3$ (left), $\ell = 5$ (middle), and $\ell = 10$ (right), power of the top $k = 3$ conditional, $\tau = 0.5$ truncated, and original Fisher's combination test for data drawn from $N(\mu, I_{10})$ with $\mu_1 = \dots = \mu_\ell$ varying according to the x-axis and $\mu_{\ell+1} = \dots = \mu_n = -2$. Power results from an average over $N = 10^4$ trials, bands denote one standard error (barely visible), and the level $\alpha = 0.1$ is denoted by the dashed line.

5.1 Simultaneous Inference on the Top k

Rather than doing inference on just the most promising effect, we may want to do simultaneous inference on the top k most promising effects. Written in terms of p-values, it is easy to give a conditional inferential procedure that does just that.

Corollary 8 (Testing the top k winners). *Suppose that p_i are n independent and selectively dominant p-values under the nulls $H_{0,i}$ and let $H_{0,(j)}$ denote the null corresponding to the j th smallest p-value. Rejecting all $H_{0,(j)}$ for which $p_{(j)} \leq \alpha_k p_{(k+1)}$ controls FWER at level α conditional on the indices of the smallest k p-values (and therefore also marginally).*

If the $k+1$ smallest p-value provides close to no evidence against the null (i.e., $p_{(k)} \approx 1$), then Corollary 8 corresponds to applying Sidak's procedure to the smallest k p-values, totally ignoring that there were n p-values to start. On the other hand, if $p_{(j)}$ not $1/\alpha_k$ times smaller than $p_{(k+1)}$, Corollary 8 will fail to reject $H_{0,(j)}$. In the parametric setting of Section 3.1.3, we get simultaneously valid conditional LCBs for the top k winning parameters (parameters corresponding to the k smallest p-values) by inverting Corollary 8's tests.

Similarly, it is straightforward to do simultaneous hybrid inference on the top k parameters.

Corollary 9 (Hybrid testing the top k winners). *Suppose that p_i are n independent and selectively dominant p-values under the nulls $H_{0,i}$. As shorthand, let $H_{0,(j)}$ denote the null corresponding to the j th smallest p-value. Fixing some $\beta < \alpha_k$, rejecting all $H_{0,(j)}$ for which*

$$p_{(j)} \leq \frac{\alpha_k - \beta}{1 - \beta} p_{(k+1)} + \left(1 - \frac{\alpha_k - \beta}{1 - \beta}\right) \beta_n \quad (41)$$

controls FWER at level α .

Like before, simultaneous hybrid inference maintains some benefits of conditional simultaneous inference while never doing worse than level β classical inference (where now β is at most α_k). Again, in Section 3.1.3's parametric setting, we can get simultaneously valid hybrid LCBs for the top k winning parameters by inverting Corollary 9's tests.

5.2 Adaptive Versions of Fisher's Combination Test

If we are just interested in verifying the existence of a signal, we can try and combine inferences on the top k effects rather than trying to simultaneous inference on them individually. This is made possible by Corollary 10, which gives a conditional version of Fisher's combination test. Recall that Fisher's original combination test considers n independent p-values p_i for the nulls $H_{0,i}$ and rejects the global null $\cap_{i=1}^n H_{i,0}$ when the test statistic $-2 \sum_{i=1}^n \log(p_i)$ is at least as large as the $1 - \alpha$ quantile of the χ_{2n}^2 distribution.

Corollary 10. Suppose that p_i are n independent and selectively dominant p -values under the nulls $H_{0,i}$ and let $H_{0,(j)}$ denote the null corresponding to the j th smallest p -value. Rejecting the data-dependent global null $\cap_{j=1}^k H_{0,(j)}$ (and therefore also the global null $\cap_{i=1}^n H_{0,i}$) when

$$-2 \sum_{j=1}^k \log(p_{(j)}/p_{(k+1)}) \geq \text{Quantile}(1 - \alpha, \chi_{2k}^2) \quad (42)$$

controls Type I error at level α conditional on the indicies of the smallest k p -values (and therefore also marginally).

Examining (42), we see that when the $(k+1)$ st p -value provides essentially no evidence against the null (so $p_{(k+1)} \approx 1$), running Corollary 10's test is like running Fisher's test using just the top k p -values, completely ignoring that we selected from n of them. In this case, our test statistic will have an essentially identical value to Fisher's original test statistic, but the critical value required for rejection will be much smaller. On the flipside, if many of the $p_{(j)}$ for $j \leq k$ are not sufficiently smaller than $p_{(k+1)}$, then Corollary 10's test statistic will be small and the test will fail to reject.

Another approach to improving Fisher's combination test is truncation: only add a p -value to Fisher's statistic if it is below some fixed threshold $\tau \in \mathbb{R}$ [Zaykin et al., 2002]. Past and previous work, however, only establishes validity of this test when the null p -values have exact $\text{Unif}[0, 1]$ distributions, excluding the possibility of one-sided null hypotheses [Zhang et al., 2020]. Corollary 11, however, gives a generalized version of Fisher's truncated combination test that is still valid whenever the p -values are independent and selectively dominant.

Corollary 11. Suppose that p_i are n independent and selectively dominant p -values under the nulls $H_{0,i}$ and fix n thresholds $\tau_i \in [0, 1]$. Letting $j \in J$ denote the random set of indices for which $p_j \leq \tau_j$, rejecting the data-dependent global null $\cap_{j \in J} H_{0,j}$ (and therefore also the global null $\cap_{i=1}^n H_{0,i}$) when

$$-2 \sum_{j=1}^k \log(p_j/\tau_j) \geq \text{Quantile}(1 - \alpha, \chi_{2|J|}^2)$$

controls type I error at level α conditional on which indicies are in J (and therefore also marginally).

If some p_j are substantially lower than their truncation point τ_j but most are above it, compared to Fisher's original combination test, Corollary 11's test will use a slightly smaller test statistic but a much smaller critical value. Hence, the truncated Fisher test is most powerful when some p -values come from strong alternatives but many p -values come from conservative nulls. (i.e, the null p -values do not have exact $\text{Unif}([0, 1])$ distributions). As such, Corollary 11 actually generalizes the truncated Fisher test to the settings where it is most useful. On the flipside, if most of the p_j are below τ_j , the truncated test statistic pays a penalty due to selection while, compared to Fisher's original test, the test's critical value remains essentially unchanged.

To illustrate the benefits and drawbacks of these methods, we display their power alongside that of Fisher's original test for a simple $n = 10$ dimensional Gaussian problem, where we sample $X \sim N(\mu, I_n)$ and use the p -values $p_i = 1 - \Phi(X_i)$ try and detect the existence of a positive mean. For $\ell \in \{3, 5, 10\}$, we vary the strength $\mu_1 = \dots = \mu_\ell > 0$ of our signals and set $\mu_{\ell+1} = \dots = \mu_n = -2$ to be conservative nulls. We do inference on the top $k = 3$ p -values using the conditional version of Fisher's method and set the truncation $\tau = 0.5$ for the truncated version (i.e., we include p_i for which $X_i > 0$). The results are displayed in Figure 4

As expected, the new methods outperform Fisher's original method when conservative nulls are present. When $\ell = 3$ and the bottom three p -values are much smaller than the rest, the conditional method does incredibly well. But its performance quickly degrades when $\ell = 5, 10$ and the fourth smallest p -value becomes often close to the bottom three. By focusing inference on just the top five p -values, the truncated method still considerably improves power when $\ell = 5$. Unsurprisingly, both methods perform worse than Fisher's original method when $\ell = 10$ and every μ_i is a signal.

5.3 Selecting by the Biggest Gap

As we saw previously, methods for doing inference using the top k effects can be powerful, but they are plagued by the question of how to choose k . These methods work well when the ratio between $p_{(k)}$ and $p_{(k+1)}$ is large, but poorly otherwise. This begs the question, why not select the p-value $p_{(j)}$ whose ratio with $p_{(j+1)}$ is the largest?

Theorem 7. *Suppose that p_i are n independent and selectively dominant p-values under the nulls $H_{0,i}$. Defining $p_{(0)} = 0$ and $p_{(n+1)} = 1$, let $G_j = \frac{p_{(j)} - p_{(j-1)}}{p_{(j+1)} - p_{(j)}}$ be the gap... Defining $W = \operatorname{argmax}_{j \in [p]} G_j$ as the index of the largest gap, the procedure that rejects $H_{0,W}$ when*

$$p_{(W)} \leq \alpha(G_{(2)}\alpha p_{(W+1)} + (1 - G_{(2)}p_{(W-1)})) \quad (43)$$

controls Type I error at level α conditional on W (and therefore also marginally)

5.4 Correcting for Publication Bias

Selectively dominant p-values also allow us to make valid inferences in the presence of publication bias. Under two models of publication bias, we use our selective dominance machinery to argue that the reader can be confident that a study's result is significant when the associated p-value is at most α^2 , even if they are worried about publication bias.

Filtered Studies: Suppose that a study tests the null H_0 at significance level α using a selectively dominant p-value p . If the study is only written-up and accepted for publication when the test is significant, then the study is essentially selecting p to use for inference when $p \leq \alpha$ (i.e., the selection function is $s(x) = I_{x \leq 1/2}$). Theorem 1 tells us that, to account for this selection, we should reject when the adjusted p-value p/α is at most α . In other words, we still control (selective) Type I error if we reject when $p \leq \alpha^2$. More generally, we show in Appendix A.9 that rejecting when $p \leq \alpha^2$ controls selective Type I error whenever the authors or publishers are using a selection function that is constant on $[0, \alpha]$ and potentially non-zero on $(\alpha, 1]$ (some non-significant p-values may be published).

p-Hacking: Rather than discard a study after observing a p-value that is more than α , a researcher may continue making tweaks to their analysis until the p-value crosses the significance threshold. This process, known as p-hacking, encompasses a vast range of techniques, and has therefore been noted to be difficult to study theoretically [Hung and Fithian, 2020]. It has been empirically well established, however, that p-hacked p-values have a left-skewed distribution under the null, i.e. p-hacked p-values can be reasonably modeled as having an increasing density under the null with support on $[0, \alpha]$ [Hung and Fithian, 2020, Simonsohn et al., 2013]. The transformed p-value p/α then has an increasing density on $[0, 1]$ under the null, and Theorem 2 therefore tells us that p/α is a (selectively dominant) p-value. Thus we can still control Type I error if we reject when $p/\alpha \leq \alpha \iff p \leq \alpha^2$.

If, in the presence of publication bias, we want to combine evidence from independent studies using selectively dominant p-values, we can run Fisher's truncated combination test (Corollary 11) with truncation point $\tau_j = \alpha$ for studies we think suffer from publication bias and $\tau_j = 1$ for those that we think do not (e.g., we may not suspect publication bias if a study reports a p-value larger than α , or if we are running a new study and are only concerned about publication bias in the previous studies). This combination test, however, requires very strong signals to reject, and, like the procedure of rejecting when $p < \alpha^2$, it will be very conservative if publication bias is not actually present.

Acknowledgements

We thank John Cherian, Yash Nair, and Will Hartog for helpful discussions.

References

- Isaiah Andrews, Toru Kitagawa, and Adam McCloskey. Inference on Winners*. *The Quarterly Journal of Economics*, 139(1):305–358, 09 2023. ISSN 0033-5533. doi: 10.1093/qje/qjad043. URL <https://doi.org/10.1093/qje/qjad043>.
- William Fithian, Dennis Sun, and Jonathan Taylor. Optimal inference after model selection, 2017. URL <https://arxiv.org/abs/1410.2597>.
- Charles J Geyer and Glen D Meeden. Fuzzy and randomized confidence intervals and p-values. *Statistical Science*, pages 358–366, 2005.
- Jesse Hemerik and Jelle Goeman. Exact testing with random permutations. *TEST*, 27(4):811–825, 2018. doi: 10.1007/s11749-017-0571-1. URL <https://doi.org/10.1007/s11749-017-0571-1>.
- Kenneth Hung and William Fithian. Rank verification for exponential families. *The Annals of Statistics*, 47(2):758 – 782, 2019. doi: 10.1214/17-AOS1634. URL <https://doi.org/10.1214/17-AOS1634>.
- Kenneth Hung and William Fithian. Statistical methods for replicability assessment. *The Annals of Applied Statistics*, 14(3):1063 – 1087, 2020. doi: 10.1214/20-AOAS1336. URL <https://doi.org/10.1214/20-AOAS1336>.
- Erich Leo Lehmann, Joseph P Romano, and George Casella. *Testing statistical hypotheses*, volume 3. Springer, 1986.
- Lihua Lei and William Fithian. AdaPT: An Interactive Procedure for Multiple Testing with Side Information. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(4):649–679, 06 2018. ISSN 1369-7412. doi: 10.1111/rssb.12274. URL <https://doi.org/10.1111/rssb.12274>.
- Ruth Marcus, Peritz Eric, and K Ruben Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3):655–660, 1976.
- Uri Simonsohn, Leif D. Nelson, and Joseph P. Simmons. P-curve: A key to the file drawer. *Cognitive Linguistics: Cognition*, 2013. URL <https://api.semanticscholar.org/CorpusID:8505270>.
- Dmitri V Zaykin, Lev A Zhivotovsky, Peter H Westfall, and Bruce S Weir. Truncated product method for combining p-values. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, 22(2):170–185, 2002.
- Hong Zhang, Tiejun Tong, John Landers, and Zheyang Wu. Tfisher: A powerful truncation and weighting procedure for combining p-values. 2020.
- Zbyněk Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967. ISSN 01621459, 1537274X. URL <http://www.jstor.org/stable/2283989>.

A Additional Derivations, Details, and Comments

A.1 Distance Between X_W and Conditional LCB

Consider observing Gaussian data $X \sim N(\mu, I_n)$ and let W and R be the indices of the winner and runner up respectively. We briefly justify that the distance between X_W and the conditional LCB for μ_W depends only on the gap between $X_W - X_R$. Define $D = X_W - \mu_0$ to be distance between X_W and the conditional LCB $\hat{\mu}$. The conditional LCB $\hat{\mu}$ satisfies

$$\frac{p^{\hat{\mu}}(X_W)}{p^{\hat{\mu}}(X_R)} = \alpha \iff \frac{1 - \Phi(X_W - \hat{\mu})}{1 - \Phi(X_R - \hat{\mu})} = \alpha \iff \frac{1 - \Phi(D)}{1 - \Phi(D - (X_W - X_R))} = \alpha.$$

Clearly then D is a function of $X_W - X_R$.

A.2 Comparing Conditional and Sidak Global Null Testing

We consider a setting where we have n independent and selectively dominant p-values p_1, \dots, p_n that are all anti-conservative, i.e., $p_j \preceq \text{Unif}([0, 1])$. At worst, these p-values are exact uniforms (e.g., they come from the boundary of the null).

We will show that, on an event with probability at least $1 - \epsilon$, the conditional procedure, which rejects when $p_{(1)} \leq \alpha p_{(2)}$, can only reject if $p_{(1)} \leq C_\epsilon/n$ for some constant $C_\epsilon > 0$. Hence, without conservative nulls, the conditional approach behaves roughly on the same order as the classical approach (Sidak).

Letting U_1, \dots, U_n be independent $\text{Unif}([0, 1])$ random variables, two facts are clear. First that $U_{(2)} \sim \text{Beta}(2, n-1)$ has mean $\frac{2}{n+1} < \frac{2}{n}$ and standard deviation $\sqrt{\frac{2(n-1)}{(n+1)^2(n+2)}} \leq \frac{2}{n}$, and second that $p_{(2)} \preceq U_{(2)}$.

Fix any $\epsilon > 0$. We have by Chebyshev's inequality that

$$\begin{aligned} P\left(p_{(2)} \leq \frac{2}{n}(1 + \epsilon^{-\frac{1}{2}})\right) &\geq P(p_{(2)} \leq E[U_{(2)}] + \sqrt{\text{Var}(U_{(2)})}/\sqrt{\epsilon}) \\ &\geq P(U_{(2)} \leq E[U_{(2)}] + \sqrt{\text{Var}(U_{(2)})}/\sqrt{\epsilon}) \\ &\geq P(|U_{(2)} - E[U_{(2)}]| < \sqrt{\text{Var}(U_{(2)})}/\sqrt{\epsilon}) \\ &\geq 1 - \epsilon \end{aligned}$$

Define $C_\epsilon = 2(1 + \epsilon^{-\frac{1}{2}})\alpha$ and the event $A_\epsilon = \{p_{(2)} \leq \frac{2}{n}(1 + \epsilon^{-\frac{1}{2}})\}$. The event A_ϵ has probability at least $1 - \epsilon$, and on this event a the conditional procedure never rejects when the procedure $p_{(1)} > C_\epsilon/n$, completing our claim.

In other words, the conditional procedure can only reject when $p_{(1)} \leq C_\epsilon/n$ (except for a small probability event) and hence suffers the same curse of dimensionality as the classical method:

$$P(p_{(1)} \leq \alpha p_{(2)} \text{ and } p_{(1)} > C_\epsilon/n) \leq P(A_\epsilon^c) \leq \epsilon.$$

A.3 Distance Between X_W and Hybrid LCB

Consider observing Gaussian data $X \sim N(\mu, I_n)$ and let W and R be the indices of the winner and runner up respectively. We briefly justify that the distance between X_W and the hybrid LCB for μ_W depends only on the gap between $X_W - X_R$ and the dimension n . Define $D = X_W - \mu_0$ to be distance between X_W and the hybrid LCB $\hat{\mu}$. The hybrid LCB $\hat{\mu}$ satisfies

$$\begin{aligned} p^{\hat{\mu}}(X_W) &= \frac{\alpha - \beta}{1 - \beta} p^{\hat{\mu}}(X_R) + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_n \\ 1 - \Phi(X_W - \hat{\mu}) &= \frac{\alpha - \beta}{1 - \beta} (1 - \Phi(X_R - \hat{\mu})) + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_n \\ \iff 1 - \Phi(D) &= \frac{\alpha - \beta}{1 - \beta} (1 - \Phi(D - (X_W - X_R))) + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_n. \end{aligned}$$

Clearly then D is a function of $X_W - X_R$ and n .

A.4 The p-Value Viewpoint of Conditional Inference

Considering data $X \sim N(\mu, I_n)$ with unknown mean μ and the p-values $p_i^{\mu_0} = 1 - \Phi(X_i - \mu_0)$, we want to characterize when the conditional LCB is at least $\mu_0 \in \mathbb{R}$. This happens exactly when μ_0 is not included in the set (12). Examining (10), (11), and (12), this happens when $X_W - \mu_0$ is at least as large as the $1 - \alpha$ quantile Q of a standard normal truncated to be larger than $X_R - \mu_0$. This quantile satisfies

$$\alpha = \frac{1 - \Phi(Q)}{1 - \Phi(X_R - \mu_0)}.$$

Solving for Q gives $Q = \Phi^{-1}(1 - \alpha(1 - \Phi(X_R - \mu_0)))$, meaning we reject exactly when

$$\begin{aligned} X_W - \mu_0 \geq \Phi^{-1}(1 - \alpha(1 - \Phi(X_R - \mu_0))) &\iff 1 - \Phi(X_W - \mu_0) \leq \alpha(1 - \Phi(X_R - \mu_0)) \\ &\iff p_{(1)}^{\mu_0} \leq \alpha p_{(2)}^{\mu_0}. \end{aligned}$$

A.5 The p-Value Viewpoint of Hybrid Inference

Considering data $X \sim N(\mu, I_n)$ with unknown mean μ and the p-values $p_i^{\mu_0} = 1 - \Phi(X_i - \mu_0)$, we want to characterize when the hybrid LCB is at least $\mu_0 \in \mathbb{R}$. Examining (16), (17), and (18), we can consider two cases to figure out when this happens.

Case One - $X_R - \mu_0 \geq z_{1-\beta_n}$: If $X_R - \mu_0 \geq z_{1-\beta_n}$, then $q_{\frac{1-\alpha}{1-\beta}}^h(W, X_{-W}, \mu_0) = -\infty$, so μ_0 cannot be in (18). This case happens precisely when

$$X_R - \mu_0 \geq z_{1-\beta_n} \iff 1 - \Phi(X_R - \mu_0) \leq 1 - \Phi(z_{1-\beta_n}) \iff p_{(2)}^{\mu_0} \leq \beta_n.$$

Case Two - $X_R - \mu_0 < z_{1-\beta_n}$: If $X_R - \mu_0 < z_{1-\beta_n}$, then μ_0 is not in (18) exactly when $X_W - \mu_0$ is at least as large as the $\frac{1-\alpha}{1-\beta}$ quantile Q of a standard normal truncated to be larger than $X_R - \mu_0$ but smaller than $z_{1-\beta_n}$. This quantile satisfies

$$\frac{\alpha - \beta}{1 - \beta} = \frac{\Phi(z_{1-\beta_n}) - \Phi(Q)}{\Phi(z_{1-\beta_n}) - \Phi(X_R - \mu_0)} = \frac{1 - \beta_n - \Phi(Q)}{1 - \beta_n - \Phi(X_R - \mu_0)}$$

Solving for Q gives

$$Q = \Phi^{-1} \left(1 - \beta_n - \frac{\alpha - \beta}{1 - \beta} (1 - \beta_n - \Phi(X_R - \mu_0)) \right) = \Phi^{-1} \left(\left(1 - \frac{\alpha - \beta}{1 - \beta} \right) (1 - \beta_n) + \frac{\alpha - \beta}{1 - \beta} \Phi(X_R - \mu_0) \right),$$

so we reject exactly when

$$\begin{aligned} X_W - \mu_0 \geq \Phi^{-1} \left(\left(1 - \frac{\alpha - \beta}{1 - \beta} \right) (1 - \beta_n) + \frac{\alpha - \beta}{1 - \beta} \Phi(X_R - \mu_0) \right) \\ \iff 1 - \Phi(X_W - \mu_0) \leq \left(1 - \frac{\alpha - \beta}{1 - \beta} \right) \beta_n + \frac{\alpha - \beta}{1 - \beta} (1 - \Phi(X_R - \mu_0)) \\ \iff p_{(1)}^{\mu_0} \leq \frac{\alpha - \beta}{1 - \beta} p_{(2)}^{\mu_0} + \left(1 - \frac{\alpha - \beta}{1 - \beta} \right) \beta_n \end{aligned}$$

It turns out we can combine these two cases. Because $p_{(1)}^{\mu_0} \leq p_{(2)}^{\mu_0}$, the fact that $p_{(2)}^{\mu_0} \leq \beta_n$ in Case One implies that $p_{(1)}^{\mu_0} \leq \beta_n$ also. Therefore in Case One, $p_{(1)}^{\mu_0}$ must be strictly smaller than a mixture of $p_{(2)}^{\mu_0}$ and β_n :

$$p_{(1)}^{\mu_0} \leq \frac{\alpha - \beta}{1 - \beta} p_{(2)}^{\mu_0} + \left(1 - \frac{\alpha - \beta}{1 - \beta} \right) \beta_n. \quad (44)$$

Therefore, if we reject according to (44), we will always reject in Case One, and we will reject at the appropriate times in Case Two.

A.6 Comparing Hybrid Inference to the Union Bound

As discussed earlier the hybrid cutoff (21) is strictly larger than the union bound cutoff (22) when $p_{(1)} > \beta_n$. Thus, both procedures reject when $p_{(1)} \leq \beta_n$. When $p_{(1)} > \beta_n$, the hybrid procedure rejects but the union bound does not whenever

$$p_{(1)} \in \left((\alpha - \beta)p_{(2)}, \frac{(\alpha - \beta)}{1 - \beta} p_{(2)} + \left(1 - \frac{(\alpha - \beta)}{1 - \beta} \right) \beta_n \right]$$

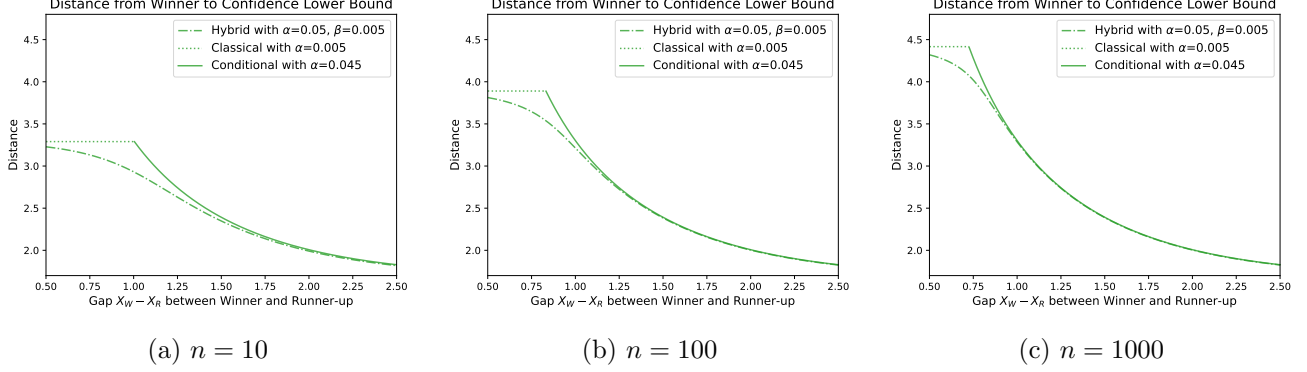


Figure 5: For $n = 10$ (left), $n = 100$ (middle), and $n = 1000$ (right) the distance between the hybrid LCB to the winner (dash-dot line) and union bound LCB to the winner (dotted and solid line) with $\alpha = 0.05$ and $\beta = 0.005$ plotted as a function of the gap between the winning and runner-up observation.

When $p_{(2)} = 1$ and $n = 1$, the length of the interval is β , which is the largest it can possibly be. Thus, in the case where we can have additional rejections, the hybrid cutoff is never more than β plus the union bound cutoff. Andrews et al. [2023] suggests taking $\beta = \alpha/10$, so when $\alpha = 0.05$, for example, $\beta = 0.005$ is quite small.

Still, this is not a precise statement about power gain. The computations required to compute the power gain analytically are messy, so instead we gauge the power gain via simulation. We sample data $X \sim N(\mu, I_n)$ for $n = 10$ and attempt to reject the winning null $H_W : \mu_W \leq 0$ where $W = \operatorname{argmin}_{i \in [n]} 1 - \Phi(X_i)$ is the index of the smallest p-value $p_i = 1 - \Phi(X_i)$.

We choose $n = 10$ because it is a reasonably small dimension size where one may apply hybrid inference (e.g., the main example from Andrews et al. [2023] has $n = 13$). Let R denote the index of the runner-up (second largest observation). For the dimensions $n = 10, 100, 1000$, Figure 5 compares the distance from the winner X_W the hybrid LCB and the union bound LCB. As illustrated in the plot, the (small) benefit of hybrid inference dissipates as the dimension of the problem increases. This is because conditioning on the good event has less and less of an effect as n grows.

Our simulation results indicate that hybrid inference typically results in a fairly small power gain. We consider two simulated settings:

Needle in a haystack: First, we consider a needle in the haystack setting where $\mu_1 > 0$ and all the other μ_i for $i \neq 1$ are set to μ_2 . We try $\mu_2 = -2, 0, 2$. The power comparison is plotted in Figure 6. Whenever we truly have a needle in the haystack problem, i.e., $\mu_1 > \mu_2$, hybrid inference results in essentially no power gain. The only setting where we see some gain (up to around 0.05) is when $\mu_2 > \mu_1$. In this setting we actually have a dense alternative (many small signals). We expect the top two p-values to be close to each other, so conditional methods should perform poorly. The union bound approach indeed performs essentially identically to the level β classical test (not pictured). Hybrid, however, manages to eke out some additional power. Both methods pale in comparison to the level α classical test however, which would achieve power > 0.95 throughout the whole plot (not pictured). For various values of σ_1 and σ_2 which we assume are known, we also tried re-running the experiments with $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_i \sim N(\mu_2, \sigma_2^2)$ when $i > 1$. The results were not appreciably different.

Two possible signals: Seeing as the hybrid and union bound approaches both reject based on the winning and running-up p-value, we ran a simulation for all pairs $\mu_1, \mu_2 \in \{-3, -2.9, \dots, 2.9, 3\}$ with $\mu_1 > \mu_2$ and $\mu_1 > 0$. We forcibly set $\mu_i = -\infty$ for $i > 2$. For each setting we ran $N = 10000$ to get an empirical estimate of power for each method. Across the 1492 simulated settings, the median empirical power increase from hybrid was ≈ 0.003 , the 90th percentile empirical power increase was ≈ 0.023 and the maximum empirical power increase was ≈ 0.042 . As the results indicate, the power increases from hybrid were negligible for most settings. We also re-ran the same simulations but sampled $X_1 \sim N(\mu, \sigma_1^2)$ and $X_2 \sim N(\mu, \sigma_2^2)$ for various values of σ_1 and σ_2 , which we assume are known. The results were not appreciably different, and, if

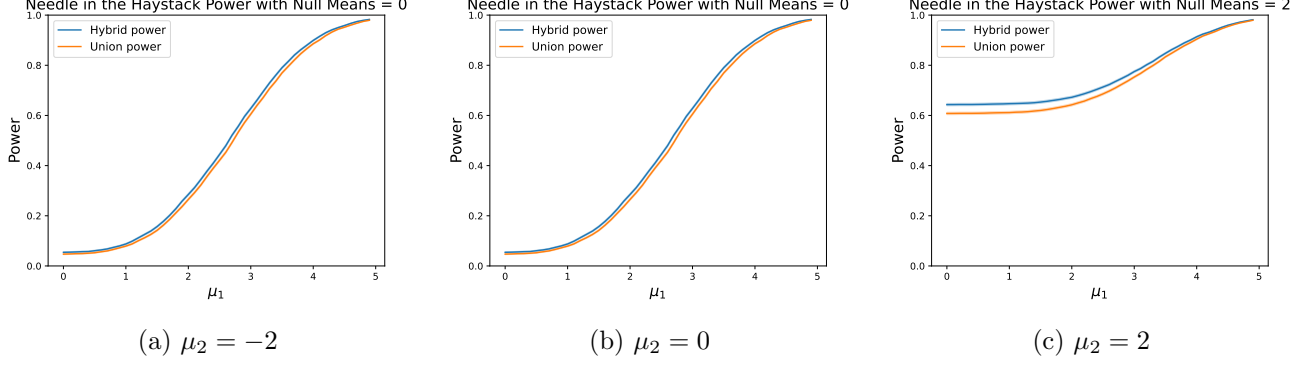


Figure 6: For $\mu_2 = 0$ (left), $\mu_2 = -0.5$ (middle), and $\mu_2 = -1$ (right) the empirical power over $N = 10^4$ trials of the hybrid inference approach versus the union bound approach for the needle in the haystack alternative. One standard error bands are also plotted. For the most part, they are so small that they are hardly visible.

anything, the difference in power was notably smaller for some values of σ_1 and σ_2 .

A.7 p-value Adjustment for Exponential Families

Suppose p is a p-value for the null H_0 that is selectively dominant given Z and we select p to use for inference according to the selection function

$$s(x, z) = \begin{cases} 1 & \text{if } x < q^+(z), \\ \frac{1}{N(z)} & \text{if } x \in [q^+(z), q(z)], \\ 0 & \text{otherwise,} \end{cases}$$

Then the adjusted p-value (4) is given by

$$p_{adj} = \frac{\int_0^p s(x, Z) dx}{\int_0^1 s(x, Z) dx} = \begin{cases} \frac{q^+(Z) + \frac{1}{N(Z)}(q(Z) - q^+(Z))}{q^+(Z) + \frac{1}{N(Z)}(q(Z) - q^+(Z))} & \text{if } p < q^+(Z), \\ \frac{q^+(Z) + \frac{1}{N(Z)}(p - q^+(Z))}{q^+(Z) + \frac{1}{N(Z)}(q(Z) - q^+(Z))} & \text{if } p \in [q^+(Z), q(Z)], \end{cases}$$

which can be re-written as

$$p_{adj} = \frac{p - (1 - \frac{1}{N(Z)})(p - q^+(Z))_+}{q^+(Z) + \frac{1}{N(Z)}(q(Z) - q^+(Z))}.$$

This is sufficient to imply the claim from the main text.

A.8 Smallest p-value and the Winner

We give a quick example showing that, in parametric settings, we cannot always use the smallest p-value to determine the winner. Suppose that $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ with the μ_i unknown but σ_i known. The UMP p-value for testing the null $H_{0,i}^{\mu_0} : \mu_i \leq \mu_0$ is given by $p_i^{\mu_0} = 1 - \Phi(\frac{X_i - \mu_0}{\sigma})$. Suppose that $\sigma_1 = 1$, $\sigma_2 = 3$ and we observe $X_1 = 1$ and $X_2 = 2$. Then

$$p_1^0 = 1 - \Phi(1) < 1 - \Phi(2/3) = p_2^0$$

but

$$p_1^1 = 1 - \Phi(0) > 1 - \Phi(1/3) = p_2^0.$$

A.9 Comparing Selection Functions

Consider a selectively dominant p-value p for the null H_0 that is valid conditional on Z , and two binary selection variables $S_1 \in 0, 1$ and $S_2 \in 0, 1$ whose joint relationships with p are governed by two different selection functions:

$$s_1(x, z) = P(S_1 = 1 | p = x, Z = z),$$

$$s_2(x, z) = P(S_2 = 1 | p = x, Z = z)$$

Sometimes, if the selection function $s_1(x, z)$ is harsher than the other $s_2(x, z)$ then using Theorem 1's adjusted p-value under $s_1(x, z)$ maintains error control even when p is selected according to $s_2(x, z)$. In particular, whenever

$$p_{adj,2} = \frac{\int_0^p s_2(x, Z) dx}{\int_0^1 s_2(x, Z) dx} \leq \frac{\int_0^p s_2(x, Z) dx}{\int_0^1 s_2(x, Z) dx} = p_{adj,1}$$

we have from Theorem 1 that

$$P_{H_0}(p_{adj,1} \leq \alpha \mid S_2 = 1) \leq P_{H_0}(p_{adj,2} \leq \alpha \mid S_2 = 1) \leq \alpha.$$

One example that is relevant to our publication bias application is where $s_1(x) = I_{x \leq \alpha}$ and $s_2(x)$ is any function that takes some constant value $c \in [0, 1]$ (we make no assumptions about what values $s_2(x)$ takes outside of $(0, \alpha]$). The adjusted p-value $p_{adj,1} = p/\alpha$ remains unchanged if we take $s_1(x) = cI_{x \leq \alpha}$, so instead (scaling the selection function by a positive constant does not change the adjustment), so instead imagine working with this selection function. Then we see that

$$\begin{aligned} \frac{\int_0^p s_2(x, Z) dx}{\int_0^1 s_2(x, Z) dx} \leq \frac{\int_0^p s_1(x, Z) dx}{\int_0^1 s_1(x, Z) dx} &\iff \frac{\int_0^1 s_2(x, Z) dx}{\int_0^p s_2(x, Z) dx} \geq \frac{\int_0^1 s_1(x, Z) dx}{\int_0^p s_1(x, Z) dx} \\ &\iff 1 + \frac{\int_p^1 s_2(x, Z) dx}{\int_0^p s_2(x, Z) dx} \geq 1 + \frac{\int_p^1 s_1(x, Z) dx}{\int_0^p s_1(x, Z) dx} \\ &\iff \frac{\int_p^1 s_2(x, Z) dx}{\int_0^p s_2(x, Z) dx} \geq \frac{\int_p^1 s_1(x, Z) dx}{\int_0^p s_1(x, Z) dx} \end{aligned}$$

Examine the last inequality. When $p > \alpha$ the numerator of the right-hand side of is 0, so it must be true. If $p \leq \alpha$, then the denominators of both sides of the last inequality are equal but the numerator of the left-hand side is at least that of the numerator of the right-hand side, which is at most $c(\alpha - p)$. Therefore, the entire string of inequalities, including the first one, is true, implying that $p_{adj,2} \leq p_{adj,1}$.

Now we briefly consider the case that $c = 0$, so $s_2(x)$ is 0 on the interval $[0, \alpha]$. It is clear for p such that $\int_0^p s_2(x) dx = 0$ that the first inequality is true (the remaining inequalities are no longer well defined since they require dividing by zero). If p is such that $\int_0^p s_2(x) dx > 0$, then we must have $p > \alpha$, so again the last inequality is true by the same reasoning as above (now the remaining inequalities are well defined). Again, we find that $p_{adj,2} \leq p_{adj,1}$.

B Selecting Multiple p-Values for Inference

In this appendix, we generalize our selective dominance framework to allow us to select multiple p-values for inference. We only do this generalization

Suppose we have n independent and selectively dominant p-values for the nulls $H_{0,i}$. We imagine conditioning on some collection of them, which, without loss of generality, we can take to be $Z = (p_{k+1}, \dots, p_n)$. Note that, due to independence, conditioning on Z does not change the distribution of p_j . Thus, the p_j remain p-values under the nulls $H_{0,j}$ that are independent. Now, for $1 \leq j \leq k$, we consider k binary selection random variables $S_j \in \{0, 1\}$, where $S_j = 1$ when p_j is selected. The relationship between p_j , Z , and S_j is governed by the selection function

$$s_j(x, z) = p(S_j = 1 | p_j = x, Z = z)$$

Supposing that U is a uniform random variable that has a uniform distribution given Z , we can imagine selecting U using the same selection functions. I.e., we can imagine that a binary selection variable $S_j \in \{0, 1\}$ whose joint distribution with U is governed by

$$P(S_j = 1 | U = x, Z = z) = s_j(x, z)$$

Then the machinery from Section 2 tells us that

$$p_{adj,j} = F_{U|Z, S_j=1}(p_j) = \frac{\int_0^{p_j} s_j(x, Z) dx}{\int_0^1 s_j(x, Z) dx}$$

is p-value (it stochastically dominates the uniform distribution under the null) conditional on Z and selection.

We will further assume that the selection happens independently, i.e.,

$$P(S_1 = 1, \dots, S_k = 1 | p_1, \dots, p_k, Z) = \prod_{j=1}^k P(S_j = 1 | p_j, Z)$$

By taking expectations conditional on Z with respect to both sides, we find that the S_j are independent given Z :

$$\begin{aligned} P(S_1 = 1, \dots, S_k = 1 | Z) &= E[P(S_1 = 1, \dots, S_k = 1 | p_1, \dots, p_k, Z) | Z] \\ &= E \left[\prod_{j=1}^k P(S_j = 1 | p_j, Z) \middle| Z \right] \\ &= \prod_{j=1}^k E[P(S_j = 1 | p_j, Z)] \\ &= \prod_{j=1}^k P(S_j = 1 | Z) \end{aligned}$$

where we have used that the p_j are conditionally independent given Z to move the expectation inside the product. Finally, conditional on Z and all the selections $S_j = 1$, the adjusted p-values $p_{adj,j}$ are all independent of one another.

$$P(p_{adj,1} \in A_1, \dots, p_{adj,k} \in A_k | Z, S_1 = 1, \dots, S_k = 1) = \prod_{j=1}^k P(p_{adj,j} \in A_j | Z, S_j = 1)$$

This can be confirmed via Bayes rule:

$$\begin{aligned} &P(p_{adj,1} \in A_1, \dots, p_{adj,k} \in A_k | Z, S_1 = 1, \dots, S_k = 1) \\ &= \frac{p(p_{adj,1} \in A_1, \dots, p_{adj,k} \in A_k, |Z) P(S_1 = 1, \dots, S_k = 1 | Z, p_{adj,1} \in A_1, \dots, p_{adj,k} \in A_k)}{P(S_1 = 1, \dots, S_k = 1 | Z)} \\ &= \frac{a + b}{c} \end{aligned}$$

C Selective Dominance and One-Sided Testing

In this appendix, we establish the selective dominance property for UMP p-values in MLR families and UMPU p-values in exponential families. We also show that in these cases, the adjusted p-value from Theorem 1 is monotone in the parameter, under suitable conditions. Throughout the appendix, we draw from the discussion and proof strategy in Appendix B.1 of [Lei and Fithian \[2018\]](#).

C.1 MLR Families

We consider a parametric family P_θ that has density $p_\theta(x)$ with respect to some carrier measure μ . These densities are such that, for any $\theta < \theta'$, the likelihood ratio $p_{\theta'}(x)/p_\theta(x)$ is a finite and non-decreasing function of some real-valued function $T(x)$. We will suppose the likelihood ratio is always finite.

Recall that for a testing problem, the critical function $\phi(x)$ (see [Lehmann et al., 1986, Section 3.1]) tells us the probability of rejecting the null having observed data x , so $\phi(X) = P(\text{reject } H_0 | X)$. From Theorem 3.4.1 of Lehmann et al. [1986] we know the test governed by the critical function

$$\phi(x) = \begin{cases} 1 & \text{if } T(X) > C \\ \gamma & \text{if } T(X) = C \\ 0 & \text{otherwise} \end{cases} \quad (45)$$

is UMP for testing $H_0 : \theta \leq \theta_0$ against the alternatives $H_a : \theta > \theta_0$ so long C and γ satisfy

$$P_{\theta_0}(T(X) > C) + \gamma P_{\theta_0}(T(X) = C) = \alpha. \quad (46)$$

Denote the left-continuous survival function of $T(X)$ and its righthand limit under P_{θ_0} as

$$G(t) = P_{\theta_0}(T(X) \geq t) \quad G^+(t) = \lim_{u \downarrow t} G(u).$$

Since $G(t)$ is a monotone non-increasing function, we can also define its generalized inverse

$$G^{-1}(z) = \sup\{t : G(t) \geq z\},$$

We know that $G(C) = G(G^{-1}(\alpha)) \geq \alpha$ by how G^{-1} is defined. If $G^+(C) > \alpha$ then we would be able to find some $t^+ > C$ such that $G(t^+) > \alpha$, contradicting that $C = G^{-1}(\alpha)$. Thus we must have $G^+(C) \leq \alpha$, finishing the proof.

$$G^{-1}(z) = \inf\{t : G(t) \leq z\},$$

Lemma C.1 gives a natural way to set C and γ in Equation (45) to get an UMP test.

Lemma C.1 (An UMP test). *Adopting the convention that $0/0 = 0$, taking $C = G^{-1}(\alpha)$ and $\gamma = (\alpha - G^+(C))/(G(C) - G^+(C))$ in Equation (45) gives an UMP test.*

Proof. At continuity points t of $G(\cdot)$, we have $G(G^{-1}(t)) = t$ and also $P(T(X) = t) = 0$. Thus, if $C = G^{-1}(\alpha)$ is a continuity point of $G(\cdot)$, then $G^+(C) = P_{\theta_0}(T(X) > C) = P_{\theta_0}(T(X) \geq C) = G(C) = \alpha$ and $\gamma = 0$, so the constraint (46) is immediately satisfied.

If $C = G^{-1}(\alpha)$ is not a continuity point of $G(\cdot)$, then $G(C) - G^+(C) > 0$, so γ is well defined in this case and the constraint (46) is also immediately satisfied. To ensure that we still have a valid test, however, we need $\gamma \in [0, 1]$. This is true so long as $\alpha \in [G^+(C), G(C)]$. We know that $G(t) \leq \alpha$ for any $t > C$, so $G^+(C) \leq \alpha$. If $G(C) < \alpha$, then we could find some $t^- < C$ such that $G(t^-) < \alpha$ by right-continuity, but this would contradict that $C = G^{-1}(\alpha)$, finishing the proof. \square

Letting $U_{aux} \sim \text{Unif}([0, 1])$ be auxiliary randomness that is independent of X , a simple way to instantiate the test from Lemma C.1 is to reject whenever $T(X) > C$ or when $T(X) = C$ and $U_{aux} \leq \gamma$. Lemma C.2 explains how this is the same as rejecting when the p-value, termed as a fuzzy p-value in Geyer and Meeden [2005],

$$p = G^+(T(X)) + U_{aux}(G(T(X)) - G^+(T(X))) \quad (47)$$

is at most α .

Lemma C.2 (Fuzzy p-value is UMP). *Rejecting $H_0 : \theta \leq \theta_0$ when the fuzzy p-value (47) is at most α instantiates the test from Lemma C.1, and is therefore UMP.*

Proof. We rewrite

$$p = (1 - U_{aux})G^+(T(X)) + U_{aux}G(T(X))$$

and consider four cases.

- If $t < G^{-1}(\alpha)$ then we can find some $t^+ > t$ such that $G(t^+) > \alpha$. Thus $G(t) \geq G^+(t) > \alpha$. So $p > \alpha$ whenever $T(X) < G^{-1}(\alpha)$
- If $t = G^{-1}(\alpha)$ and $G^{-1}(\alpha)$ is a continuity point of $G(\cdot)$, then $G^+(t) = G(t) = \alpha$. Thus in this case $p \leq \alpha$ whenever $T(X) = G^{-1}(\alpha)$ and $U_{aux} \leq \frac{\alpha - G^+(G^{-1}(\alpha))}{G(G^{-1}(\alpha)) - G^+(G^{-1}(\alpha))} = \infty$.
- If $t = G^{-1}(\alpha)$ and $G^{-1}(\alpha)$ is not a continuity point of $G(\cdot)$, then we must have that $G(t) - G^+(t) > 0$. Also by right continuity we have $G(t) \geq \alpha$ and by how $G^{-1}(\cdot)$ is defined we have $G^+(t) \leq \alpha$. In this case $p \leq \alpha$, also whenever $T(X) = G^{-1}(\alpha)$ and $U_{aux} \leq \frac{\alpha - G^+(G^{-1}(\alpha))}{G(G^{-1}(\alpha)) - G^+(G^{-1}(\alpha))}$.
- If $t > G^{-1}(\alpha)$ then $G^+(t) \leq G(t) \leq \alpha$. So $p \leq \alpha$ whenever $T(X) > G^{-1}(\alpha)$.

This implies the following set equalities:

$$\begin{aligned} \{p \leq \alpha\} &= \{T(X) > G^{-1}(\alpha)\} \cup \left\{ T(X) = G^{-1}(\alpha), U_{aux} \leq \frac{\alpha - G^+(G^{-1}(\alpha))}{G(G^{-1}(\alpha)) - G^+(G^{-1}(\alpha))} \right\} \\ &= \{T(X) > C\} \cup \{T(X) = C, U_{aux} \leq \gamma\} \end{aligned}$$

□

Lemma C.3 shows that $p \sim \text{Unif}([0, 1])$ under P_{θ_0} , which will be useful for us later.

Lemma C.3 (Fuzzy p-value is uniform at null boundary). *Under P_{θ_0} , the p-value (47) has a $\text{Unif}([0, 1])$ distribution.*

Proof. Using the set equality from Lemma C.2 but replacing α with $z \in (0, 1)$, we find

$$\begin{aligned} P_{\theta_0}(G^+(T(X)) + U_{aux}(G(T(X)) - G^+(T(X))) &\leq z) \\ &= P_{\theta_0}(T(X) > G^{-1}(z)) + P_{\theta_0}\left(T(X) = G^{-1}(z), U \leq \frac{z - G^+(G^{-1}(z))}{G(G^{-1}(z)) - G^+(G^{-1}(z))}\right) \\ &= P_{\theta_0}(T(X) > G^{-1}(z)) + P_{\theta_0}(T(X) = G^{-1}(z))P_{\theta_0}\left(U \leq \frac{z - G^+(G^{-1}(z))}{G(G^{-1}(z)) - G^+(G^{-1}(z))}\right) \\ &= G^+(G^{-1}(z)) + (G(G^{-1}(z)) - G^+(G^{-1}(z))) \cdot \frac{z - G^+(G^{-1}(z))}{G(G^{-1}(z)) - G^+(G^{-1}(z))} \\ &= z. \end{aligned}$$

□

Now we can show that p is a selectively dominant p-value for testing the null $H_0 : \theta \leq \theta_0$. To start, we prove some facts that allow us to relate the distribution of $T(X)$ under P_{θ_0} and under P_θ for $\theta \leq \theta_0$.

Lemma C.4. *Consider some $\theta \leq \theta_0$ and let $g_\theta(T(x)) = p_\theta(x)/p_{\theta_0}(x)$ be the non-increasing likelihood ratio and*

$$\nu(A) = \int I(T(X) \in A) p_{\theta_0}(x) \mu(dx)$$

be the measure of $T(X)$ under $X \sim P_{\theta_0}$. Then

$$P_\theta(T(X) \in A) = \int_A g_\theta(t) \nu(dt)$$

Proof. This follows from an application of the “standard machine” (see the discussion after Equation 42 in [Lei and Fithian \[2018\]](#)). □

Lemma C.5. *If ω denotes the product measure of $T(X)$ and Lesbesgue measure on $[0, 1]$, i.e., the distribution of $(T(X), U_{aux})$ under P_{θ_0} , then*

$$P_{\theta_0}((T(X), U) \in B) = \int_B g_{\theta}(t) \omega(dt, du)$$

$$\nu(A) = \int I(T(X) \in A) p_{\theta_0}(x) \mu(dx)$$

be the measure of $T(X)$ under $X \sim P_{\theta_0}$. Then

$$P_{\theta}(T(X) \in A) = \int_A g_{\theta}(t) \nu(dt)$$

Proof. This follows from a standard application of the $\pi - \lambda$ theorem. \square

With these lemmas under our belt, we can show Proposition C.1. Since $p \sim_{P_{\theta_0}} \text{Unif}([0, 1])$, this proposition is sufficient to imply selective dominance.

Proposition C.1. *Consider some $\theta \leq \theta_0$ and any selection function $s(x)$ for which p has positive probability of selection under both θ and θ_0 . Then*

$$P_{\theta}(p \leq z | S = 1) \leq P_{\theta_0}(p \leq z | S = 1)$$

Proof. For $\theta \leq \theta_0$ define $g_{\theta}(T(x)) = p_{\theta}(x)/p_{\theta_0}(x)$ to be the non-increasing likelihood ratio. Define the mapping

$$m(t, u) = G^+(t) + u(G(t) - G^+(t))$$

so that $p = m(T(X), U_{aux})$, and also define the reparametrized the selection function

$$s(t, u) = s(m(t, u)).$$

We will denote the pre-image of $m(\cdot, \cdot)$ by $m^{-1}(\cdot)$.

It is straightforward to show that via Bayes rule that

$$P_{\theta}(p \in A | S = 1) = \frac{\int_{m^{-1}(A)} g_{\theta}(t) s(t, u) \omega(dt, du)}{\int_{m^{-1}([0, 1])} g_{\theta}(t) s(t, u) \omega(dt, du)}.$$

If z is such that $P_{\theta}(p \leq z | S = 1) = 0$ then the desired inequality is trivial. To handle the non-trivial case, note that if $(t, u) \in m^{-1}([0, z])$ then $t \geq G^{-1}(z)$, and if $(t, u) \in m^{-1}((z, 1])$ then $t \leq G^{-1}(z)$. Also, because m is an invertible mapping, $m^{-1}([0, z])$ and $m^{-1}((z, 1])$ are disjoint. Thus,

$$\begin{aligned} \frac{1}{P_{\theta}(p \leq z | S = 1)} &= \frac{\int_{m^{-1}([0, 1])} g_{\theta}(t) s(t, u) \omega(dt, du)}{\int_{m^{-1}([0, z])} g_{\theta}(t) s(t, u) \omega(dt, du)} \\ &= \frac{\int_{m^{-1}([0, z])} g_{\theta}(t) s(t, u) \omega(dt, du) + \int_{m^{-1}((z, 1])} g_{\theta}(t) s(t, u) \omega(dt, du)}{\int_{m^{-1}([0, z])} g_{\theta}(t) s(t, u) \omega(dt, du)} \\ &= 1 + \frac{\int_{m^{-1}((z, 1])} g_{\theta}(t) s(t, u) \omega(dt, du)}{\int_{m^{-1}([0, z])} g_{\theta}(t) s(t, u) \omega(dt, du)} \\ &\geq 1 + \frac{g_{\theta}(G^{-1}(z)) \int_{m^{-1}((z, 1])} s(t, u) \omega(dt, du)}{g_{\theta}(G^{-1}(z)) \int_{m^{-1}([0, z])} s(t, u) \omega(dt, du)} \\ &= 1 + \frac{\int_{m^{-1}((z, 1])} s(t, u) \omega(dt, du)}{\int_{m^{-1}([0, z])} s(t, u) \omega(dt, du)} \\ &= \frac{1}{P_{\theta_0}(p \leq z | S = 1)}, \end{aligned}$$

where at the end, we have noted that $g_{\theta_0}(t) = 1$ everywhere. This is sufficient to imply the claim. \square

Proposition C.2. Let $s_\theta(x)$ be a selection function such that for $\theta \leq \theta_0$ the ratio $s_{\theta_0}(x)/s_\theta(x)$ is non-decreasing in x .

that is monotone non-decreasing in θ for every $x \in [0, 1]$. If p_{adj}^θ the adjustment from $F_{U|S_\theta=1}(p^\theta)$ then p^θ is monotone non-decreasing increasing in θ .

Proof. Fix a t and u and let $A_{t,u} = \{(x, y) : x > t \text{ or } x = t \text{ and } y \leq u\}$. Then it suffices to show that

$$\frac{\int_{A_{t,u}} g_\theta(t) s_\theta(t, u) \omega(dt, du)}{\int_{A_{t,u}} s_\theta(t, u) \omega(dt, du)}$$

is monotone.

$$\frac{\int_{A_{t,u}} g_\theta(t) s_\theta(t, u) \omega(dt, du)}{\int_{A_{t,u}} s_\theta(t, u) \omega(dt, du)}$$

$$\begin{aligned} \frac{p_{(1)}^\theta}{p_{(2)}^\theta} &= \frac{\int_{A_{t_1, u_1}} g_\theta(t) \omega(dt, du)}{\int_{A_{t_2, u_2}} g_\theta(t) \omega(dt, du)} \\ &= 1 + \frac{\int_{A_{t_1, u_1}} g_\theta(t) \omega(dt, du)}{\int_{A_{t_2, u_2} \setminus A_{t_1, u_1}} g_\theta(t) \omega(dt, du)} \\ &\leq 1 + \frac{g_\theta(t_1) \int_{A_{t_1, u_1}} \omega(dt, du)}{g_\theta(t_1) \int_{A_{t_2, u_2} \setminus A_{t_1, u_1}} \omega(dt, du)} \\ &= 1 + \frac{\int_{A_{t_1, u_1}} \omega(dt, du)}{\int_{A_{t_2, u_2} \setminus A_{t_1, u_1}} \omega(dt, du)} \\ &= \frac{p_{(1)}^{\theta_0}}{p_{(2)}^{\theta_0}} \end{aligned}$$

□

C.2 Proof of Corollary 2, Corollary 4, and Corollary 6

In this appendix, we prove the validity of all our LCBs for MLR families. We consider a MLR family P_θ parameterized by $\theta \in \mathbb{R}$ as in ???. Suppose we observe n independent samples $X_i \sim P_{\theta_i}$, and are imagining using the UMP p-values $p_i^{\theta_0}$ for the test $H_{0,i} : \theta_i \leq \theta_0$ as given in (9).

Proving that our provided confidence regions have at least $1 - \alpha$ coverage is straightforward, but arguing that the resulting confidence regions are actually LCBs is more tricky. We first need a simple analysis lemma.

Lemma C.6. If $f(x) \geq 0$ is non-increasing real-valued function in $x \in \mathbb{R}$ and $a \geq b$ with $a, b \in (0, 1)$, then

$$\frac{ah(x) + (1 - a)}{bh(x) + (1 - b)} \tag{48}$$

is also a non-increasing function in x .

Proof. Considering some $y \leq z$, rearranging things shows that

$$\begin{aligned} \frac{ah(y) + (1 - a)}{bh(y) + (1 - b)} \geq \frac{ah(z) + (1 - a)}{bh(z) + (1 - b)} &\iff \frac{(a - b)(h(y) - h(z))}{(ah(x) + (1 - a))(bh(x) + (1 - b))} \geq 0 \\ &\iff (a - b)(h(y) - h(z)) \geq 0 \end{aligned}$$

and the right-hand inequality is clearly true.

□

And then another key lemma.

Lemma C.7. *Both $p_{(1)}^{\theta_0}$ and $p_{(1)}^{\theta_0}/p_{(2)}^{\theta_0}$ are non-decreasing in θ_0 .*

Proof. For a t and u , let $A_{t,u} = \{(x, y) : x > t \text{ or } x = t \text{ and } y \leq u\}$. Then

$$\begin{aligned} \frac{p_{(1)}^{\theta}}{p_{(2)}^{\theta}} &= \frac{\int_{A_{t_1, u_1}} g_{\theta}(t) \omega(dt, du)}{\int_{A_{t_2, u_2}} g_{\theta}(t) \omega(dt, du)} \\ &= 1 + \frac{\int_{A_{t_1, u_1}} g_{\theta}(t) \omega(dt, du)}{\int_{A_{t_2, u_2} \setminus A_{t_1, u_1}} g_{\theta}(t) \omega(dt, du)} \\ &\leq 1 + \frac{g_{\theta}(t_1) \int_{A_{t_1, u_1}} \omega(dt, du)}{g_{\theta}(t_1) \int_{A_{t_2, u_2} \setminus A_{t_1, u_1}} \omega(dt, du)} \\ &= 1 + \frac{\int_{A_{t_1, u_1}} \omega(dt, du)}{\int_{A_{t_2, u_2} \setminus A_{t_1, u_1}} \omega(dt, du)} \\ &= \frac{p_{(1)}^{\theta_0}}{p_{(2)}^{\theta_0}} \end{aligned}$$

□

C.3 Exponential Families

D Proofs and Derivations

D.1 Proof of Theorem 1

Recall we are considering a selection function such that the probability that U is selected $P(S = 1) = \int_0^1 s(x) dx$ is positive. The CDF of U given selection is continuous because it cannot have any point masses:

$$P(U = x | S = 1) = \frac{P(U = x, S = 1)}{P(S = 1)} \leq \frac{P(U = x)}{P(S = 1)} = 0.$$

Therefore, defining

$$F_{U|S=1}^{-1}(t) = \inf\{x : F_{U|S=1}(x) > t\}$$

continuity implies that $F_{U|S=1}(F_{U|S=1}^{-1}(t)) = t$ and $F_{U|S=1}(x) \leq t \iff x \leq F_{U|S=1}^{-1}(t)$. Then

$$P_{H_0}(F_{U|S=1}(p) \leq t | S = 1) = P_{H_0}(p \leq F_{U|S=1}^{-1}(t) | S = 1) \leq P(U \leq F_{U|S=1}^{-1}(t) | S = 1) = P(F_{U|S=1}(U) \leq t | S = 1)$$

where we have used that $p|S = 1 \succeq_{H_0} U|S = 1$ to get the middle inequality. Finally

$$P(F_{U|S=1}(U) \leq t | S = 1) = P(U \leq F_{U|S=1}^{-1}(t) | S = 1) = F_{U|S=1}(F_{U|S=1}^{-1}(t)) = t,$$

so $F_{U|S=1}(U) | S = 1 \sim \text{Unif}([0, 1])$ which finishes the proof.

D.2 Proof of Theorem 2

Let f be the density of p under a distribution in the null H_0 . We start by showing that, if f is non-decreasing, then $p|S = 1$ dominates $U|S = 1$. Fixing a selection function $s(x)$, it suffices to show that for any $t \in [0, 1]$,

$$\begin{aligned} P(p \leq t | S = 1) &\leq P(U \leq t | S = 1) \\ \iff \frac{\int_0^t s(x) f(x) dx}{\int_0^1 s(x) f(x) dx} &\leq \frac{\int_0^t s(x) dx}{\int_0^1 s(x) dx} \end{aligned}$$

If $P(p \leq t | S = 1)$ is zero then this trivially holds. Otherwise $P(p \leq t | S = 1) = \int_0^t s(x)f(x)dx > 0$ and we see that,

$$\begin{aligned} \frac{\int_0^1 s(x)f(x)dx}{\int_0^t s(x)f(x)dx} &= 1 + \frac{\int_t^1 s(x)f(x)dx}{\int_0^t s(x)f(x)dx} \\ &\geq 1 + \frac{f(t) \int_t^1 s(x)dx}{f(t) \int_0^t s(x)dx} \\ &= 1 + \frac{\int_t^1 s(x)dx}{\int_0^t s(x)dx} \\ &= \frac{\int_0^1 s(x)dx}{\int_0^t s(x)dx}, \end{aligned}$$

which is sufficient to imply the claim.

Now assuming that f is continuous and non-decreasing, we can show that p is not selectively dominant. In general, it suffices for there to be two points $y_1 < y_2$ such that f is strictly larger in a neighborhood around y_1 than in a neighborhood around y_2 , where these neighborhoods are disjoint. In particular for $\epsilon > 0$ let $N_\epsilon(y) = (y - \epsilon, y + \epsilon)$ be a ball around y . Then we need there to be $y_1, y_2, \epsilon > 0$, and some $\eta > 0$ such that, for all $z_1 \in N_\epsilon(y_1)$ and $z_2 \in N_\epsilon(y_2)$, $z_1 < z_2$ but $f(z_2) + \eta < f(z_1)$. If f is continuous and not non-decreasing, then this must be true. First define $B_{high} = \inf\{f(z_1) : z_1 \in N_\epsilon(y_1)\}$ and $B_{low} = \sup\{f(z_2) : z_2 \in N_\epsilon(y_2)\}$ so $B_{high} > B_{low}$. Then consider the selection function

$$s(x) = \begin{cases} 1 & \text{if } x \in N_\epsilon(y_1) \cup N_\epsilon(y_2) \\ 0 & \text{otherwise} \end{cases}$$

and let t be a value such that $t > z_1$ for all $z_1 \in N_\epsilon(y_1)$ and $t < z_2$ for all $z_2 \in N_\epsilon(y_2)$. Trivially,

$$P(U \leq t | S = 1) = \frac{1}{2}.$$

But the fact that

$$\begin{aligned} \frac{1}{P(p \leq t | S = 1)} &= \frac{\int_{y_1-\epsilon}^{y_1+\epsilon} f(x)dx + \int_{y_2-\epsilon}^{y_2+\epsilon} f(x)dx}{\int_{y_1-\epsilon}^{y_1+\epsilon} f(x)dx} \\ &= 1 + \frac{\int_{y_2-\epsilon}^{y_2+\epsilon} f(x)dx}{\int_{y_1-\epsilon}^{y_1+\epsilon} f(x)dx} \\ &\leq 1 + \frac{2\epsilon B_{low}}{2\epsilon B_{high}} \\ &< 2 \end{aligned}$$

implies that $P(p \leq t | S = 1) > \frac{1}{2}$, which means that p cannot be selectively dominant.

Now we can give proofs of each corollary.

D.2.1 Proof of Corollary 2

Proof of Corollary 4

Proof of Corollary 6

D.3 Proof of Corollary 1

It suffices to argue that closing Sidak's rejects $H_{0,(k)}$ if and only if $p_{(j)} \leq \alpha_{n-j+1}$ for every $1 \leq j \leq k$. For a subset $I \subseteq [p]$, Sidak's global null test rejects $H_{0,I}$ when the smallest p-value in I is at most $\alpha_{|I|}$.

Necessity: For $1 \leq j \leq k$, let I_{n-j+1} be the size $n-j+1$ subset that excludes the $j-1$ smallest p-values for (when $j=1$ then $I=[p]$). This subset includes the index of the k th smallest p-value, so we must reject $H_{0,I}$ to reject $H_{0,(k)}$. It rejects exactly when $p_{(j)} \leq \alpha_{n-j+1}$, so our conditions are necessary.

Sufficiency: Consider a size $n-j+1$ size subset I that has the k th smallest p-value. The smallest p-value in this subset is the ℓ th smallest p-value for some $\ell \leq k$ and also $\ell \leq j$. We reject $H_{I,0}$ because

$$p_{(\ell)} \leq \alpha_{n-\ell+1} \leq \alpha_{n-j+1} = \alpha_I.$$

D.4 Proof of Corollary 3

It suffices to argue that closing our conditional global null testing procedure rejects $H_{0,(k)}$ if and only if $p_{(j)} \leq \alpha p_{(j+1)}$ for every $1 \leq j \leq k$. We will define $p_{(n+1)} = 1$ and, for subsets $I \subseteq [p]$ of size one, we define the conditional procedure to reject the global null $H_{I,0}$ when the lone p-value is at most α . For subsets I of size strictly more than one, the procedure rejects the global null $H_{I,0}$ when the smallest p-value in I is at most α times the second smallest p-value in I .

Necessity: For $1 \leq j \leq k$, let I_{n-j+1} be the size $n-j+1$ subset that excludes the $j-1$ smallest p-values for (when $j=1$ then $I=[p]$). This subset includes the index of the k th smallest p-value, so we must reject $H_{0,I}$ to reject $H_{0,(k)}$. It rejects exactly when $p_{(j)} \leq \alpha p_{(j+1)}$, so our conditions are necessary.

Sufficiency: Consider a subset I that contains the index of the k th smallest p-value. If it is size one, then we reject because

$$p_{(k)} \leq \alpha p_{(k+1)} \leq \alpha$$

If it is of size at least two, suppose the smallest p-value in the set is then the ℓ th smallest p-value for $\ell \leq k$ and the second smallest p-value in I is m th smallest p-value for some $m > \ell$. We will reject $H_{0,I}$ because

$$p_{(\ell)} \leq \alpha p_{(\ell+1)} \leq \alpha p_{(m)}.$$

D.5 Proof of Theorem 5

Suppose we have n independent and selectively dominant p-values p_i for the null hypotheses $H_{0,i}$. We restrict our attention to $j \in \mathcal{J}$ for which p_j has positive probability of being the smallest. Suppose we use p_j to test $H_{0,j}$ only when we observe that p_j is strictly larger than β_n but still the smallest of all the p-values. We can apply Section 2's framework with $p = p_j$, $Z = p_{-j}$ and the selection function $s(x, z) = I_{\beta_n < p_j < \min_{i \neq j} p_i}$. It is straightforward to see that our adjusted p-value p_{adj} is $(p_j - \beta_n) / (\min_{i \neq j} p_i - \beta_n)$, and Theorem 1 therefore tells us that

$$P_{H_{0,j}} \left(\frac{p_j - \beta_n}{\min_{i \neq j} p_i - \beta_n} \leq \frac{\alpha - \beta}{1 - \beta} \mid p_{-j}, S = 1 \right) \leq \frac{\alpha - \beta}{1 - \beta}.$$

Re-arranging things we get

$$P_{H_{0,j}} \left(p_j \leq \frac{\alpha - \beta}{1 - \beta} \min_{i \neq j} p_i + \left(1 - \frac{\alpha - \beta}{1 - \beta} \right) \beta_n \mid p_{-j}, S = 1 \right) \leq \frac{\alpha - \beta}{1 - \beta}. \quad (49)$$

Letting W be the index of the smallest p-value, we can now prove the claim that rejecting $H_{0,W}$ when

$$p_{(1)} \leq \frac{\alpha - \beta}{1 - \alpha} p_{(2)} + \left(1 - \frac{\alpha - \beta}{1 - \alpha} \right) \beta_n$$

controls Type I error at level α . Let \widetilde{W} be the index of the smallest p-value if all the p-values are strictly larger than β_n . If some p-value is at most β_n , then force $\widetilde{W} = 0$. Let G_P be the event that a p-value corresponding to a true null is at most β_n (note the event G depends on the data generating process P). We know from Sidak's procedure (Theorem 3) that $P(G_P) \leq \beta$. Three facts are immediate:

$$G_P \subseteq \{\widetilde{W} = 0\} \implies P(\widetilde{W} = 0) \geq \beta \implies P(\widetilde{W} > 0) \leq 1 - \beta$$

$$P(\text{falsely reject } H_{0,W}, G_P, \widetilde{W} = 0) \leq P(G_P) \leq \beta,$$

$$P(\text{falsely reject } H_{0,W}, G_P^c, \widetilde{W} = 0) = 0.$$

If $H_{0,j}$ is true, the event $\widetilde{W} = j$ is the same event as selecting p_j for inference in (50), so

$$\begin{aligned} P(\text{falsely reject } H_{0,W} | \widetilde{W} = j) &= P\left(p_{(1)} \leq \frac{\alpha - \beta}{1 - \alpha} p_{(2)} + \left(1 - \frac{\alpha - \beta}{1 - \alpha}\right) \beta_n | \widetilde{W} = j\right) \\ &= P\left(p_j \leq \frac{\alpha - \beta}{1 - \alpha} \min_{i \neq j} p_i + \left(1 - \frac{\alpha - \beta}{1 - \alpha}\right) \beta_n | \widetilde{W} = j\right) \\ &\leq \frac{\alpha - \beta}{1 - \alpha}, \end{aligned}$$

and if $H_{0,j}$ is not true then trivially $P(\text{falsely reject } H_{0,W} | \widetilde{W} = j) = 0 \leq \frac{\alpha - \beta}{1 - \alpha}$. Our result then follows from law of total probability:

$$\begin{aligned} &P(\text{falsely reject } H_{0,W}) \\ &= P(\text{falsely reject } H_{0,W}, \widetilde{W} = 0, G_P) + P(\text{falsely reject } H_{0,W}, \widetilde{W} = 0, G_P^c) \\ &\quad + \sum_{j \in \mathcal{J}} P(\text{falsely reject } H_{0,W} | \widetilde{W} = j) P(\widetilde{W} = j) \\ &\leq \beta + \frac{\alpha - \beta}{1 - \beta} \sum_{j \in \mathcal{J}} P(\widetilde{W} = j) \\ &= \beta + \frac{\alpha - \beta}{1 - \beta} P(\widetilde{W} > 0) \\ &\leq \alpha. \end{aligned}$$

D.6 Proof of Corollary 5

It suffices to argue that closing our hybrid global null testing procedure rejects $H_{0,(k)}$ if and only if

$$p_{(j)} \leq \frac{\alpha - \beta}{1 - \beta} p_{(j+1)} + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_{n-j+1}$$

for every $1 \leq j \leq k$. We will define $p_{(n+1)} = \alpha$ so that the right-hand side of the above equals α when $j = n$. Correspondingly, for subsets $I \subseteq [p]$ of size one, we define the hybrid procedure to reject the global null $H_{I,0}$ when the lone p-value is at most α . For subsets I of size strictly more than one, supposing that the smallest p-value in I is the ℓ th smallest p-value and the second smallest p-value in I is the m th smallest p-value, the hybrid procedure rejects the global null $H_{I,0}$ when

$$p_{(\ell)} \leq \frac{\alpha - \beta}{1 - \beta} p_{(m)} + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_{|I|}.$$

Necessity: For $1 \leq j \leq k$, let I_{n-j+1} be the size $n - j + 1$ subset that excludes the $j - 1$ smallest p-values (when $j = 1$ then $I = [p]$). This subset includes the index of the k th smallest p-value, so we must reject $H_{0,I}$ to reject $H_{0,(k)}$. It rejects exactly when

$$p_{(j)} \leq \frac{\alpha - \beta}{1 - \beta} p_{(j+1)} + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_{n-j+1}$$

so our conditions are necessary.

Sufficiency: Consider a subset I that contains the index of the k th smallest p-value. If it is size-one, then we reject because

$$p_{(k)} \leq \frac{\alpha - \beta}{1 - \beta} p_{(k+1)} + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_{n-k+1} \leq \frac{\alpha - \beta}{1 - \beta} + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta = \alpha.$$

Now suppose that I is size $n - j + 1$. Its smallest p-value is the ℓ th smallest p-value for some $\ell \leq k$ and $\ell \leq j$, and its second smallest p-value is the m th smallest p-value for some $m > \ell$. We reject because

$$\begin{aligned} p_{(\ell)} &\leq \frac{\alpha - \beta}{1 - \beta} p_{(\ell+1)} + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_{n-\ell+1} \\ &\leq \frac{\alpha - \beta}{1 - \beta} p_{(m)} + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_{n-j+1} \\ &= \frac{\alpha - \beta}{1 - \beta} p_{(m)} + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_{|I|}. \end{aligned}$$

D.7 Proof of Theorem 7

Suppose we have n independent and selectively dominant p-values p_i for the null hypotheses $H_{0,i}$. We restrict our attention to $j \in \mathcal{J}$ for which their is positive probability that the ratio of p_j to the next largest p-value is the smallest (with this ratio defined to be just the p-value itself for the largest p-value). For $x \in \mathbb{R}$ and $z \in \mathbb{R}^{n-1}$, define $y(x, z) \in \mathbb{R}^{n-1}$ to be the sorted vector $y_1(x, z) \leq \dots \leq y_n(x, z) \leq y_{n+1}(x, z) = 1$ whose entries are composed of x , the entries of z , and ends with a one. Let $k(x, z)$ be the smallest index such that $y_{k(x, z)}(x, z) = x$. Then, we can apply Section 2's framework with $p = p_j$, $Z = p_{-j}$ and the selection function

$$s(x, z) = \begin{cases} 1 & \text{if } \frac{y_{k(x, z)}(x, z)}{y_{k(x, z)+1}(x, z)} \leq \min_{j \in [n]: j \neq k(x, z)} \frac{y_j}{y_{j+1}} \\ 0, & \text{otherwise} \end{cases}$$

where we take $0/0 = 1$. An equivalent characterization of this selection function is, defining the sorted vector $\tilde{y}(z) \in \mathbb{R}^n$ be the entries of z with a one tacked on at the end. Let I be the set of indices for which $\tilde{y}_i(z)/\tilde{y}_{i+1}(z)$ is minimized. Then $s(x, z)$ is one when $x \in \bigcup_{i \in I} [\tilde{y}_i(z)]$

$s(x, z) = I_{\beta_n < p_j < \min_{i \neq j} p_i}$. It is straightforward to see that our adjusted p-value p_{adj} is $(p_j - \beta_n)/(\min_{i \neq j} p_i - \beta_n)$, and Theorem 1 therefore tells us that

$$P_{H_{0,j}} \left(\frac{p_j - \beta_n}{\min_{i \neq j} p_i - \beta_n} \leq \frac{\alpha - \beta}{1 - \beta} \mid p_{-j}, S = 1 \right) \leq \frac{\alpha - \beta}{1 - \beta}.$$

Re-arranging things we get

$$P_{H_{0,j}} \left(p_j \leq \frac{\alpha - \beta}{1 - \beta} \min_{i \neq j} p_i + \left(1 - \frac{\alpha - \beta}{1 - \beta}\right) \beta_n \mid p_{-j}, S = 1 \right) \leq \frac{\alpha - \beta}{1 - \beta}. \quad (50)$$

Letting W be the index of the smallest p-value, we can now prove the claim that rejecting $H_{0,W}$ when

$$p_{(1)} \leq \frac{\alpha - \beta}{1 - \alpha} p_{(2)} + \left(1 - \frac{\alpha - \beta}{1 - \alpha}\right) \beta_n$$

controls Type I error at level α . Let \widetilde{W} be the index of the smallest p-value if all the p-values are strictly larger than β_n . If some p-value is at most β_n , then force $\widetilde{W} = 0$. Let G_P be the event that a p-value corresponding to a true null is at most β_n (note the event G depends on the data generating process P). We know from Sidak's procedure (Theorem 3) that $P(G_P) \leq \beta$. Three facts are immediate: