Submitted to the Annals of Statistics

SUPPLEMENTARY MATERIAL TO 'ROBUST INFERENCE WITH KNOCKOFFS'

BY RINA FOYGEL BARBER*, EMMANUEL J. CANDÈS†, AND RICHARD J. SAMWORTH‡

The University of Chicago, Stanford University† and University of Cambridge‡*

APPENDIX B: ADDITIONAL PROOFS

B.1. Proof of Theorem 3. First, we will show that our statement can be reduced to a binary hypothesis testing problem. We will work under the global null hypothesis where $Y \perp\!\!\!\perp X$, and our test will be constructed independently of Y. More formally, let $P_{Y|X}$ be any fixed distribution, e.g. $\mathcal{N}(0,1)$. Since all features are null, this means that the false discovery proportion is 1 whenever $\widehat{\mathcal{S}}(\mathbf{X},\mathbf{Y}) \neq \emptyset$, that is,

$$\mathrm{FDR}\big(\widehat{\mathcal{S}}\big) = \mathbb{P}\left\{\widehat{\mathcal{S}}(\mathbf{X},\mathbf{Y}) \neq \emptyset\right\}.$$

Therefore, in order to prove the theorem, it is sufficient to construct a *binary* test $\psi(\mathbf{X}) \in \{0,1\}$ such that

(B.1)

$$\mathbb{P}_{\mathbf{X}_{i,*} \overset{\text{iid}}{\sim} P^\star_X}\left\{\psi(\mathbf{X}) = 1\right\} \geq q\big(1 + c(1-e^{-\epsilon})\big), \quad \mathbb{P}_{\mathbf{X}_{i,*} \overset{\text{iid}}{\sim} P_X}\left\{\psi(\mathbf{X}) = 1\right\} = q,$$

i.e. a test ψ that has better-than-random performance for testing whether the conditional distribution of X_j is given by $P^\star_j$ or P_j. Once ψ is constructed, then this is sufficient for the FDR result, e.g. setting

$$\widehat{\mathcal{S}}(\mathbf{X},\mathbf{Y}) = \begin{cases} \{j\}, & \psi(\mathbf{X}) = 1,\\ \emptyset, & \psi(\mathbf{X}) = 0.\end{cases}$$

Note that, by the well-known equivalence between total variation distance and hypothesis testing [Lehmann and Romano, 2008], the existence of such a test ψ is essentially equivalent to proving a lower bound on

$$\mathrm{d_{TV}}\left((P^\star_X)^{\otimes n}, \big(P_X\big)^{\otimes n}\right)$$

uniformly over all distributions P_X whose jth conditional is P_j. In fact, our ψ will be given by a randomized procedure (to be fully formal, we can use the independent

imsart-aos ver. 2012/08/31 file: rk_supp_rev2.tex date: April 18, 2019

random vector $\mathbf{Y}$ as a source of randomness, if needed). First, we draw $\widetilde{\mathbf{X}} \mid \mathbf{X}$, independently of $\mathbf{Y}$ and drawn from the rule $P_{\widetilde{X}|X}$ as specified in the theorem, and independently we also draw $B \sim \text{Bernoulli}(2q)$ and $B' \sim \text{Bernoulli}(q)$. Next, defining $\widehat{\text{KL}}_j$ as in (13), we let

$$\psi(\mathbf{X}, \widetilde{\mathbf{X}}, B, B') = \mathbb{1}\left\{B = 1 \text{ and } \widehat{\text{KL}}_j > 0\right\} + \mathbb{1}\left\{B' = 1 \text{ and } \widehat{\text{KL}}_j = 0\right\}.$$

Clearly, by definition of B and B', we have

$$\mathbb{P}\left\{\psi(\mathbf{X}, \widetilde{\mathbf{X}}, B, B') = 1\right\} = 2q \cdot \mathbb{P}\left\{\widehat{\text{KL}}_j > 0\right\} + q \cdot \mathbb{P}\left\{\widehat{\text{KL}}_j = 0\right\}, \tag{B.2}$$

where $\mathbb{P}\left\{\widehat{\text{KL}}_j > 0\right\}$ and $\mathbb{P}\left\{\widehat{\text{KL}}_j = 0\right\}$ are taken with respect to the joint distribution of $(\mathbf{X}, \widetilde{\mathbf{X}})$.

Next, we check that the test ψ satisfies the properties (B.1), as required for the FDR bounds in this theorem. We first prove the second bound in (B.1). Suppose $\mathbf{X}_{i,*} \overset{\text{iid}}{\sim} P_X$—that is, P_j is indeed the correct conditional distribution for $X_j \mid X_{-j}$. The knockoff generating mechanism $P_{\widetilde{X}|X}$ was defined to satisfy pairwise exchangeability with respect to P_j (5), meaning that $\mathbf{X}_j$ and $\widetilde{\mathbf{X}}_j$ are exchangeable conditional on the other variables in this scenario. Examining the form of $\widehat{\text{KL}}_j$, we see that swapping $\mathbf{X}_j$ and $\widetilde{\mathbf{X}}_j$ has the effect of changing the sign of $\widehat{\text{KL}}_j$. The exchangeability of the pair $(\mathbf{X}_j, \widetilde{\mathbf{X}}_j)$ implies that the distribution of $\widehat{\text{KL}}_j$ is symmetric around zero, and so under $(\mathbf{X}_{i,*}, \widetilde{\mathbf{X}}_{i,*}) \overset{\text{iid}}{\sim} P_X \times P_{\widetilde{X}|X}$,

$$\mathbb{P}\left\{\widehat{\text{KL}}_j > 0\right\} + 0.5 \cdot \mathbb{P}\left\{\widehat{\text{KL}}_j = 0\right\} = 0.5.$$

Checking (B.2), this proves that $\mathbb{P}_{\mathbf{X}_{i,*} \overset{\text{iid}}{\sim} P_X}\left\{\psi(\mathbf{X}) = 1\right\} = q$, which ensures FDR control for the case that the estimated conditional P_j is in fact correct.

Finally we turn to the first part of (B.1), where now we assume that $(\mathbf{X}_{i,*}, \widetilde{\mathbf{X}}_{i,*}) \overset{\text{iid}}{\sim} P^\star_X \times P_{\widetilde{X}|X}$. From this point on, we will condition on the observed values of $\mathbf{X}_{-j}$ and $\widetilde{\mathbf{X}}_{-j}$. By assumption in the theorem, under this distribution we have $\mathbb{P}\left\{\widehat{\text{KL}}_j \geq \epsilon\right\} \geq c$. As in the proof of Lemma 2, we consider the unordered pair $\{\mathbf{X}_j, \widetilde{\mathbf{X}}_j\}$—that is, we see the two vectors $\mathbf{X}_j$ and $\widetilde{\mathbf{X}}_j$ but do not know which is which. Note that, with this information, we are able to compute $\left|\widehat{\text{KL}}_j\right|$ but not $\text{sign}(\widehat{\text{KL}}_j)$. Without loss of generality, we can label the unordered pair of feature vectors $\{\mathbf{X}_j, \widetilde{\mathbf{X}}_j\}$, as $\mathbf{X}_j^{(0)}$ and $\mathbf{X}_j^{(1)}$, such that

- if $\mathbf{X}_j = \mathbf{X}_j^{(0)}$ and $\widetilde{\mathbf{X}}_j = \mathbf{X}_j^{(1)}$, then $\widehat{\text{KL}}_j \geq 0$;

- if $\mathbf{X}_j = \mathbf{X}_j^{(1)}$ and $\widetilde{\mathbf{X}}_j = \mathbf{X}_j^{(0)}$, then $\widehat{\mathrm{KL}}_j \leq 0$.

Define $C = \mathrm{sign}(\widehat{\mathrm{KL}}_j)$, so that $\widehat{\mathrm{KL}}_j = C \cdot \left|\widehat{\mathrm{KL}}_j\right|$. By definition of the distribution of $(\mathbf{X}, \widetilde{\mathbf{X}})$, it follows from Lemma 1 that

$$\frac{\mathbb{P}\left\{(\mathbf{X}_j, \widetilde{\mathbf{X}}_j) = (\mathbf{X}_j^{(0)}, \mathbf{X}_j^{(1)}) \;\middle|\; \mathbf{X}_j^{(0)}, \mathbf{X}_j^{(1)}, \mathbf{X}_{-j}\right\}}{\mathbb{P}\left\{(\mathbf{X}_j, \widetilde{\mathbf{X}}_j) = (\mathbf{X}_j^{(1)}, \mathbf{X}_j^{(0)}) \;\middle|\; \mathbf{X}_j^{(0)}, \mathbf{X}_j^{(1)}, \mathbf{X}_{-j}\right\}} \\ = \prod_i \frac{P_j^\star(\mathbf{X}_{ij}^{(0)} \mid \mathbf{X}_{i,-j}) P_j(\mathbf{X}_{ij}^{(1)} \mid \mathbf{X}_{i,-j})}{P_j(\mathbf{X}_{ij}^{(0)} \mid \mathbf{X}_{i,-j}) P_j^\star(\mathbf{X}_{ij}^{(1)} \mid \mathbf{X}_{i,-j})}.$$

In other words, if $|\widehat{\mathrm{KL}}_j| \neq 0$, then

$$\frac{\mathbb{P}\left\{C = +1 \;\middle|\; \mathbf{X}_j^{(0)}, \mathbf{X}_j^{(1)}, \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}\right\}}{\mathbb{P}\left\{C = -1 \;\middle|\; \mathbf{X}_j^{(0)}, \mathbf{X}_j^{(1)}, \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}\right\}} = \prod_i \frac{P_j^\star(\mathbf{X}_{ij}^{(0)} \mid \mathbf{X}_{i,-j}) P_j(\mathbf{X}_{ij}^{(1)} \mid \mathbf{X}_{i,-j})}{P_j(\mathbf{X}_{ij}^{(0)} \mid \mathbf{X}_{i,-j}) P_j^\star(\mathbf{X}_{ij}^{(1)} \mid \mathbf{X}_{i,-j})} \\ = \exp\left\{\left|\widehat{\mathrm{KL}}_j\right|\right\},$$

where the last step holds by our choice of which vector to label as $\mathbf{X}^{(0)}$ and which to label as $\mathbf{X}^{(1)}$.

Therefore, we can write

$$\begin{aligned} c \leq \mathbb{P}\left\{\widehat{\mathrm{KL}}_j \geq \epsilon\right\} &= \mathbb{P}\left\{C = +1 \text{ and } \left|\widehat{\mathrm{KL}}_j\right| \geq \epsilon\right\} \\ &= \mathbb{E}\left[\mathbb{P}\left\{C = +1 \;\middle|\; \mathbf{X}_j^{(0)}, \mathbf{X}_j^{(1)}, \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}\right\} \cdot \mathbb{1}\left\{\left|\widehat{\mathrm{KL}}_j\right| \geq \epsilon\right\}\right] \\ &= \mathbb{E}\left[\frac{e^{\left|\widehat{\mathrm{KL}}_j\right|}}{1 + e^{\left|\widehat{\mathrm{KL}}_j\right|}} \cdot \mathbb{1}\left\{\left|\widehat{\mathrm{KL}}_j\right| \geq \epsilon\right\}\right]. \end{aligned} \tag{B.3}$$

We can similarly calculate

$$\mathbb{P}\left\{\widehat{\mathrm{KL}}_j > 0\right\} = \mathbb{E}\left[\frac{e^{\left|\widehat{\mathrm{KL}}_j\right|}}{1 + e^{\left|\widehat{\mathrm{KL}}_j\right|}} \cdot \mathbb{1}\left\{\left|\widehat{\mathrm{KL}}_j\right| > 0\right\}\right].$$

Therefore,

$$\begin{aligned}
&\frac{1}{2}\mathbb{P}\left\{\widehat{\mathrm{KL}}_j = 0\right\} + \mathbb{P}\left\{\widehat{\mathrm{KL}}_j > 0\right\} \\
&= \mathbb{E}\left[\frac{e^0}{1+e^0}\cdot \mathbb{1}\left\{\left|\widehat{\mathrm{KL}}_j\right| = 0\right\}\right] + \mathbb{E}\left[\frac{e^{\left|\widehat{\mathrm{KL}}_j\right|}}{1+e^{\left|\widehat{\mathrm{KL}}_j\right|}}\cdot \mathbb{1}\left\{\left|\widehat{\mathrm{KL}}_j\right| > 0\right\}\right] \\
&= \mathbb{E}\left[\frac{e^{\left|\widehat{\mathrm{KL}}_j\right|}}{1+e^{\left|\widehat{\mathrm{KL}}_j\right|}}\right].
\end{aligned}$$

To continue, observe that for $t \geq 0$, $e^t/(1+e^t) \geq 1/2$. Hence,

$$\begin{aligned}
\mathbb{E}\left[\frac{e^{\left|\widehat{\mathrm{KL}}_j\right|}}{1+e^{\left|\widehat{\mathrm{KL}}_j\right|}}\right] &\geq \frac{1}{2} + \mathbb{E}\left[\left(\frac{e^{\left|\widehat{\mathrm{KL}}_j\right|}}{1+e^{\left|\widehat{\mathrm{KL}}_j\right|}} - \frac{1}{2}\right)\cdot \mathbb{1}\left\{\left|\widehat{\mathrm{KL}}_j\right| \geq \epsilon\right\}\right] \\
&\geq \frac{1}{2} + \min_{t\geq\epsilon}\frac{\frac{e^t}{1+e^t}-\frac{1}{2}}{\frac{e^t}{1+e^t}}\cdot \underbrace{\mathbb{E}\left[\frac{e^{\left|\widehat{\mathrm{KL}}_j\right|}}{1+e^{\left|\widehat{\mathrm{KL}}_j\right|}}\cdot \mathbb{1}\left\{\left|\widehat{\mathrm{KL}}_j\right| \geq \epsilon\right\}\right]}_{\geq c \text{ by (B.3)}} \\
&\geq \frac{1}{2}\left(1 + c(1-e^{-\epsilon})\right),
\end{aligned}$$

where for the last step we check that the minimum is attained at $t = \epsilon$. This proves that, when $\mathbf{X}_{i,*} \overset{\text{iid}}{\sim} P^\star_X$, we have $\psi(\mathbf{X}, \widetilde{\mathbf{X}}, B, B') = 1$ with probability at least $q\big(1 + c(1-e^{-\epsilon})\big)$, and so the first part of (B.1) is satisfied, as desired.

B.2. Proof of Lemma 3. We will in fact prove a more general result, which will be useful later on:

LEMMA B.1. *Fix any $\delta \geq 0$, and define the event*

$$\mathcal{E}_\delta = \left\{\sum_i \left[\log\left(\frac{P^\star_j(\mathbf{X}_{ij} \mid \mathbf{X}_{i,-j})P_j(\widetilde{\mathbf{X}}_{ij} \mid \mathbf{X}_{i,-j})}{P_j(\mathbf{X}_{ij} \mid \mathbf{X}_{i,-j})P^\star_j(\widetilde{\mathbf{X}}_{ij} \mid \mathbf{X}_{i,-j})}\right)\right]^2 \leq n\delta^2 \textit{ for all } j\right\}.$$

Then

$$\mathbb{P}\left\{\max_{j=1,\dots,p}\widehat{\mathrm{KL}}_j \leq \frac{n\delta^2}{2} + 2\delta\sqrt{n\log(p)}\right\} \geq 1 - \frac{1}{p} - \mathbb{P}\left\{(\mathcal{E}_\delta)^c\right\}.$$

In order to prove Lemma 3, then, we simply observe that if the universal bound (17) holds for the likelihood ratios, then the event $\mathcal{E}_\delta$ occurs with probability 1.

Now we prove the general result, Lemma B.1. Fix any j. Suppose that we condition on $\mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}$, and on the unordered pair $\{\mathbf{X}_{ij}, \widetilde{\mathbf{X}}_{ij}\} = \{a_{ij}, b_{ij}\}$ for each i—that is, after observing the unlabeled pair, we arbitrarily label them as a and b. Write $a_j = (a_{1j}, \dots, a_{nj})$ and same for b_j. Let $C_{ij} = 0$ if $a_{ij} = b_{ij}$, and otherwise let

$$C_{ij} := \begin{cases} +1, & \text{if } (\mathbf{X}_{ij}, \widetilde{\mathbf{X}}_{ij}) = (a_{ij}, b_{ij}), \\ -1, & \text{if } (\mathbf{X}_{ij}, \widetilde{\mathbf{X}}_{ij}) = (b_{ij}, a_{ij}). \end{cases}$$

Then we have

$$\begin{aligned} \widehat{\mathrm{KL}}_j &= \sum_i \log\left(\frac{P_j^\star(\mathbf{X}_{ij} \mid \mathbf{X}_{i,-j}) \cdot P_j(\widetilde{\mathbf{X}}_{ij} \mid \mathbf{X}_{i,-j})}{P_j(\mathbf{X}_{ij} \mid \mathbf{X}_{i,-j}) \cdot P_j^\star(\widetilde{\mathbf{X}}_{ij} \mid \mathbf{X}_{i,-j})}\right) \\ &= \sum_i C_{ij} \log\left(\frac{P_j^\star(a_{ij} \mid \mathbf{X}_{i,-j}) \cdot P_j(b_{ij} \mid \mathbf{X}_{i,-j})}{P_j(a_{ij} \mid \mathbf{X}_{i,-j}) \cdot P_j^\star(b_{ij} \mid \mathbf{X}_{i,-j})}\right) =: \sum_i C_{ij} \widehat{\mathrm{KL}}_{ij}. \end{aligned}$$

By Lemma 1, for each i with $a_{ij} \neq b_{ij}$ we have

$$\begin{aligned} &\frac{\mathbb{P}\left\{C_{ij} = +1 \,\middle|\, a_j, b_j, \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}\right\}}{\mathbb{P}\left\{C_{ij} = -1 \,\middle|\, a_j, b_j, \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}\right\}} \\ &\qquad = \frac{\mathbb{P}\left\{(\mathbf{X}_{ij}, \widetilde{\mathbf{X}}_{ij}) = (a_{ij}, b_{ij}) \,\middle|\, a_j, b_j, \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}\right\}}{\mathbb{P}\left\{(\mathbf{X}_{ij}, \widetilde{\mathbf{X}}_{ij}) = (b_{ij}, a_{ij}) \,\middle|\, a_j, b_j, \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}\right\}} \\ &\qquad = \frac{P_j^\star(a_{ij} \mid \mathbf{X}_{i,-j}) P_j(b_{ij} \mid \mathbf{X}_{i,-j})}{P_j(a_{ij} \mid \mathbf{X}_{i,-j}) P_j^\star(b_{ij} \mid \mathbf{X}_{i,-j})} = e^{\widehat{\mathrm{KL}}_{ij}}. \end{aligned} \tag{B.4}$$

Note that this binary outcome is independent for each i. From this point on we treat $\mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}, a_j, b_j$ as fixed (where $a_j = (a_{1j}, \dots, a_{nj})$ and same for b_j), and only the C_{ij}'s as random. Since $\widehat{\mathrm{KL}}_{ij}$ depends only on $\mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}, a_j, b_j$ (i.e. on the variables that we are conditioning on), and is therefore treated as fixed, while $|C_{ij}| \leq 1$ by definition, we see that, writing $\mu_j = \mathbb{E}\left[\widehat{\mathrm{KL}}_j \,\middle|\, \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}, a_j, b_j\right]$,

$$\mathbb{P}\left\{\widehat{\mathrm{KL}}_j - \mu_j \geq 2\sqrt{\log(p)}\sqrt{\sum_i (\widehat{\mathrm{KL}}_{ij})^2} \,\middle|\, \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}, a_j, b_j\right\} \leq \frac{1}{p^2}$$

by Hoeffding's inequality. Next we work with the conditional expectation of $\widehat{\mathrm{KL}}_j$. For any i with $a_{ij} \neq b_{ij}$, we use (B.4) to calculate

$$\left|\mathbb{E}\left[C_{ij} \,\middle|\, \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}, a_j, b_j\right]\right| = \left|\frac{e^{\widehat{\mathrm{KL}}_{ij}} - 1}{e^{\widehat{\mathrm{KL}}_{ij}} + 1}\right| \leq \frac{|\widehat{\mathrm{KL}}_{ij}|}{2}.$$

Then

$$\left|\mathbb{E}\left[\widehat{\mathrm{KL}}_j \;\middle|\; \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}, a_j, b_j\right]\right| = \left|\sum_i \mathbb{E}\left[C_{ij} \;\middle|\; \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}, a_j, b_j\right] \cdot \widehat{\mathrm{KL}}_{ij}\right| \\ \leq \frac{1}{2}\sum_i (\widehat{\mathrm{KL}}_{ij})^2.$$

Therefore, combining everything,

$$\mathbb{P}\left\{\widehat{\mathrm{KL}}_j \geq \frac{1}{2}\sum_i (\widehat{\mathrm{KL}}_{ij})^2 + 2\sqrt{\log(p)}\sqrt{\sum_i (\widehat{\mathrm{KL}}_{ij})^2} \;\middle|\; \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}, a_j, b_j\right\} \leq \frac{1}{p^2}.$$

Now, under the event $\mathcal{E}_\delta$ we must have $\sum_i (\widehat{\mathrm{KL}}_{ij})^2 \leq n\delta^2$, and so we can write

$$\mathbb{P}\left\{\widehat{\mathrm{KL}}_j \cdot \mathbb{1}\left\{\mathcal{E}_\delta\right\} \geq \frac{n\delta^2}{2} + 2\delta\sqrt{n\log(p)} \;\middle|\; \mathbf{X}_{-j}, \widetilde{\mathbf{X}}_{-j}, a_j, b_j\right\} \leq \frac{1}{p^2}.$$

Marginalizing over all the conditioned variables, and taking a union bound over all j, we have proved that

$$\mathbb{P}\left\{\max_{j=1,\dots,p} \widehat{\mathrm{KL}}_j \cdot \mathbb{1}\left\{\mathcal{E}_\delta\right\} \geq \frac{n\delta^2}{2} + 2\delta\sqrt{n\log(p)}\right\} \leq \frac{1}{p}.$$

This proves the lemma.

B.3. Proof of Lemma 4. Fix any feature index j, and consider any distribution $D^{(j)}$ on $\mathbb{R}^p$ with jth conditional equal to P_j, as defined in (19). For simplicity, from this point on, we will perform calculations treating $D^{(j)}$ as a joint density, but the result is valid without this assumption. Drawing $X \sim D^{(j)}$ and $\widetilde{X} \mid X \sim P_{\widetilde{X}|X}(\cdot|X)$, then the joint density of $(X, \widetilde{X})$ is given by

$$D^{(j)}(x) \cdot P_{\widetilde{X}|X}(\widetilde{x} \mid x) \\ = \underbrace{D^{(j)}_{-j}(x_{-j})}_{\text{Term 1}} \cdot \underbrace{\left(\frac{P_j(x_j \mid x_{-j})}{\exp\left\{-\frac{1}{2}x^\top \widetilde{\Theta} x\right\}}\right)}_{\text{Term 2}} \cdot \underbrace{\left(P_{\widetilde{X}|X}(\widetilde{x} \mid x) \cdot \exp\left\{-\frac{1}{2}x^\top \widetilde{\Theta} x\right\}\right)}_{\text{Term 3}},$$

where $D^{(j)}_{-j}$ is the marginal distribution of X_{-j} under the joint distribution $X \sim D^{(j)}$. In order to check that $(X_j, \widetilde{X}_j, X_{-j}, \widetilde{X}_{-j}) \stackrel{\mathrm{d}}{=} (\widetilde{X}_j, X_j, X_{-j}, \widetilde{X}_{-j})$ under this distribution, we therefore need to check that this joint density is exchangeable in the variables x_j and $\widetilde{x}_j$; that is, swapping x_j and $\widetilde{x}_j$ does not change the value of the joint density $D^{(j)}(x) \cdot P_{\widetilde{X}|X}(\widetilde{x} \mid x)$. We check this by considering each of the three terms separately. Term 1 clearly does not depend on either x_j or $\widetilde{x}_j$. Next, using the calculation of P_j in (19), we can simplify Term 2 to obtain

$$\text{Term 2} \propto \exp\left\{-\frac{1}{2/\widetilde{\Theta}_{jj}}\left(x_j + x_{-j}^\top \widetilde{\Theta}_{-j,j}/\widetilde{\Theta}_{jj}\right)^2 + \frac{1}{2}x^\top \widetilde{\Theta} x\right\}$$
$$= \exp\left\{\frac{1}{2}x_{-j}^\top\left(\widetilde{\Theta}_{-j,-j} - \frac{\widetilde{\Theta}_{-j,j}\widetilde{\Theta}_{-j,j}^\top}{\widetilde{\Theta}_{jj}}\right)x_{-j}\right\},$$

which also does not depend on either x_j or $\widetilde{x}_j$. Finally, Term 3 is exchangeable in the pair $x_j, \widetilde{x}_j$ by the construction of the knockoff distribution $P_{\widetilde{X}|X}$. More concretely, using the definition of $P_{\widetilde{X}|X}$ given in (18), we can calculate

Term 3

$$\propto \exp\left\{-\frac{1}{2}\big(\widetilde{x} - (\mathbf{I} - D\widetilde{\Theta})x\big)^\top (2D - D\widetilde{\Theta}D)^{-1}\big(\widetilde{x} - (\mathbf{I} - D\widetilde{\Theta})x\big) - \frac{1}{2}x^\top \widetilde{\Theta} x\right\}$$
$$= \exp\left\{-\frac{1}{2}(x + \widetilde{x})^\top (2D - D\widetilde{\Theta}D)^{-1}(x + \widetilde{x}) + x^\top D^{-1}\widetilde{x}\right\},$$

which is clearly exchangeable in the pair $x_j, \widetilde{x}_j$ (note that the exchangeability of $x_j, \widetilde{x}_j$ in the term $x^\top D^{-1}\widetilde{x}$ follows from the fact that D is a diagonal matrix).

B.4. Proof of Lemma 5. We will apply Lemma B.1 to prove this result. We first recall the conditional distributions $P_j^\star$ for the joint distribution $P_X^\star = \mathcal{N}_p(0, \Theta)^{-1}$, which can be computed as

$$P_j^\star(\cdot|x_{-j}) = \mathcal{N}\left(x_{-j}^\top\left(-\Theta_{-j,j}/\Theta_{jj}\right), 1/\Theta_{jj}\right),$$

and the conditionals P_j, calculated earlier in (19) as

$$P_j(\cdot|x_{-j}) = \mathcal{N}\left(x_{-j}^\top\left(-\widetilde{\Theta}_{-j,j}/\widetilde{\Theta}_{jj}\right), 1/\widetilde{\Theta}_{jj}\right).$$

Then we can calculate

$$\sum_i \left[\log\left(\frac{P_j^\star(\mathbf{X}_{ij} \mid \mathbf{X}_{i,-j}) P_j(\widetilde{\mathbf{X}}_{ij} \mid \mathbf{X}_{i,-j})}{P_j(\mathbf{X}_{ij} \mid \mathbf{X}_{i,-j}) P_j^\star(\widetilde{\mathbf{X}}_{ij} \mid \mathbf{X}_{i,-j})}\right)\right]^2$$
$$= \sum_i \left[-(\mathbf{X}_{ij} - \widetilde{\mathbf{X}}_{ij}) \cdot \frac{\widetilde{\Theta}_{jj} - \Theta_{jj}}{2} + \mathbf{X}_{i*}^\top (\widetilde{\Theta}_j - \Theta_j)\right]^2 \cdot \left[\mathbf{X}_{ij} - \widetilde{\mathbf{X}}_{ij}\right]^2$$
$$\leq \frac{1}{2}\sum_i \Big[\underbrace{-(\mathbf{X}_{ij} - \widetilde{\mathbf{X}}_{ij}) \cdot \frac{\widetilde{\Theta}_{jj} - \Theta_{jj}}{2} + \mathbf{X}_{i*}^\top (\widetilde{\Theta}_j - \Theta_j)}_{\sim \mathcal{N}(0, v_j^2) \text{ for each } i}\Big]^4$$
$$+ \frac{1}{2}\sum_i \Big[\underbrace{\mathbf{X}_{ij} - \widetilde{\mathbf{X}}_{ij}}_{\sim \mathcal{N}(0, w_j^2) \text{ for each } i}\Big]^4.$$

Using standard tail bounds on Gaussian and χ^2 random variables, and computing the variances v_j^2 and w_j^2, after some calculations we can show that the quantity above is bounded as

$$\sum_i \left[\log\left(\frac{P_j^\star(\mathbf{X}_{ij} \mid \mathbf{X}_{i,-j}) P_j(\widetilde{\mathbf{X}}_{ij} \mid \mathbf{X}_{i,-j})}{P_j(\mathbf{X}_{ij} \mid \mathbf{X}_{i,-j}) P_j^\star(\widetilde{\mathbf{X}}_{ij} \mid \mathbf{X}_{i,-j})}\right)\right]^2$$
$$\leq 4\left[\left(\frac{\delta_\Theta}{1-\delta_\Theta}\right)^2 + \left(\frac{\delta_\Theta}{1-\delta_\Theta}\right)^4\right] \cdot \left(\sqrt{n} + 2\sqrt{\log(np)}\right)^2,$$

with probability at least $1 - \frac{1}{p}$, and therefore, $\mathbb{P}\left\{\mathcal{E}_\delta\right\} \geq 1 - \frac{1}{p}$ when we take

$$\delta = 2\sqrt{\left(\frac{\delta_\Theta}{1-\delta_\Theta}\right)^2 + \left(\frac{\delta_\Theta}{1-\delta_\Theta}\right)^4} \cdot \left(1 + 2\sqrt{\frac{\log(np)}{n}}\right) = 2\delta_\Theta \cdot (1 + o(1)),$$

where the last step holds as long as $\frac{\log(p)}{n} = o(1)$ and $\delta_\Theta = o(1)$. Applying Lemma B.1 then proves that

$$\mathbb{P}\left\{\max_{j=1,\dots,p} \widehat{\mathrm{KL}}_j \leq \frac{n\delta^2}{2} + 2\delta\sqrt{n\log(p)}\right\} \geq 1 - \frac{2}{p}.$$

Assuming this upper bound on the $\widehat{\mathrm{KL}}_j$'s is bounded by a constant, the dominant term is therefore $2\delta\sqrt{n\log(p)}$, which proves the lemma.

B.5. Proof of Lemma 6. First, recall that $T = T(W)$ is defined as follows:

$$T = \min\left\{t \geq \epsilon(W) : \underbrace{\frac{c + \sum_{\ell=1}^p \mathbb{1}\left\{W_\ell \leq -t\right\}}{\sum_{\ell=1}^p \mathbb{1}\left\{W_\ell \geq t\right\}}}_{=: f(W,t)} \leq q\right\},$$

where $\epsilon(W) > 0$ is chosen to be the smallest magnitude of the W statistics, i.e. $\epsilon(W) = \min\{|W_\ell| : |W_\ell| > 0\}$, and where $c = 0$ for knockoff or $c = 1$ for knockoff+. Next, define

$$W^j := (W_1, \ldots, W_{j-1}, |W_j|, W_{j+1}, \ldots, W_p)$$

and similarly

$$W^k := (W_1, \ldots, W_{k-1}, |W_k|, W_{k+1}, \ldots, W_p),$$

so that $T_j = T(W^j)$ and $T_k = T(W^k)$. Note that $|W^j| = |W^k| = |W|$, and so $\epsilon(W^j) = \epsilon(W^k) = \epsilon(W)$ since $\epsilon(W)$ depends on W only through $|W|$.

Without loss of generality, assume $T_j \le T_k$, so that by assumption we have $W_j \le -T_j$ and $W_k \le -T_j$. Consider

$$f(W^k, T_j) = \frac{c + \sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^k \le -T_j\right\}}{\sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^k \ge T_j\right\}}.$$

We will next rewrite the numerator and denominator. Beginning with the numerator, we have

$$\begin{aligned}
&\sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^k \le -T_j\right\} = \sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^j \le -T_j\right\} + \\
&\mathbb{1}\left\{W_j^k \le -T_j\right\} - \mathbb{1}\left\{W_j^j \le -T_j\right\} + \mathbb{1}\left\{W_k^k \le -T_j\right\} - \mathbb{1}\left\{W_k^j \le -T_j\right\} \\
&\qquad = \sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^j \le -T_j\right\} + (1 - 0 + 0 - 1) = \sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^j \le -T_j\right\},
\end{aligned}$$

where the first step holds since W^j and W^k differ only on entries j, k, while the second step holds because we know from our assumptions and definitions that $W_j^k = W_j \le -T_j$, $W_j^j = |W_j| \ge T_j$, $W_k^k = |W_k| \ge T_j$, and $W_k^j = W_k \le -T_j$. Similarly, for the denominator, we have

$$\begin{aligned}
&\sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^k \ge T_j\right\} = \sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^j \ge T_j\right\} + \\
&\quad \mathbb{1}\left\{W_j^k \ge T_j\right\} - \mathbb{1}\left\{W_j^j \ge T_j\right\} + \mathbb{1}\left\{W_k^k \ge T_j\right\} - \mathbb{1}\left\{W_k^j \ge T_j\right\} \\
&\qquad = \sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^j \ge T_j\right\} + (0 - 1 + 1 - 0) = \sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^j \ge T_j\right\}.
\end{aligned}$$

Therefore,

$$f(W^k, T_j) = \frac{c + \sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^k \leq -T_j\right\}}{\sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^k \geq T_j\right\}} = \frac{c + \sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^j \leq -T_j\right\}}{\sum_{\ell=1}^p \mathbb{1}\left\{W_\ell^j \geq T_j\right\}}$$
$$= f(W^j, T_j) \leq q,$$

where the last step holds by definition of T_j. Therefore, since $T_j \geq \epsilon(W^j) = \epsilon(W^k)$, we see from the definition of T_k that we must have $T_k \leq T_j$. This proves that $T_j = T_k$, as desired.

REFERENCES

E.L. Lehmann and J.P. Romano. *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer New York, 2008.

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF CHICAGO
CHICAGO, IL, U.S.A.
E-MAIL: rina@uchicago.edu

DEPARTMENTS OF STATISTICS AND MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CA, U.S.A.
E-MAIL: candes@stanford.edu

STATISTICAL LABORATORY
UNIVERSITY OF CAMBRIDGE
CAMBRIDGE, U.K.
E-MAIL: r.samworth@statslab.cam.ac.uk