Supplement to "An Omnibus Nonparametric Test of Equality in Distribution for Unknown Functions"

Alexander R. Luedtke

Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

E-mail: aluedtke@fredhutch.org

Marco Carone

Department of Biostatistics, University of Washington, Seattle, WA, USA

Mark J. van der Laan

Division of Biostatistics, University of California, Berkeley, Berkeley, CA, USA

Supplementary Appendix A: Pathwise differentiability

We now review first- and second-order pathwise differentiability. We start by defining a fluctuation submodel through P_0 for which the closure of the linear span of all such scores yields the unrestricted tangent space $L_0^2(P_0)$, i.e. the set of P_0 mean zero functions in $L^2(P_0)$. Note that it is the resulting (first-order) tangent space that is important, as all differentiability properties discussed in this appendix are equivalent for any set of functions h_1 that yield the same tangent space. In the submodel that we present shortly we will restrict the score function to be uniformly bounded, but this will have no impact on the resulting differentiability properties since the closure of the linear span of these scores will yield the unrestricted tangent space. Note also that the second-order tangent space is also determined by the first-order tangent space (see Carone et al., 2014 and the references therein).

We now present the submodel that we will use in this work. Define the following fluctuation submodel through P_0:

$$dP_t(o) \triangleq \left(1 + th_1(o) + t^2 h_2(o)\right) dP_0(o),$$
$$\text{where } P_0 h_j = 0 \text{ and } \sup_{o\in\mathcal{O}} |h_j(o)| < \infty, j = 1, 2.$$

One can verify that the function h_1 is the score of the submodel at $t = 0$.

Let $\psi_t \triangleq \Psi(P_t)$. The parameter Ψ is called (first-order) pathwise differentiable at P_0 if there exists a $D_1^\Psi \in L_0^2(P_0)$ such that

$$\psi_t - \psi_0 = tP_0 D_1^\Psi h_1 + o(t).$$

We call D_1^Ψ the first-order canonical gradient of Ψ at P_0, where we note that $D_1^\Psi(O)$ is almost surely unique because $\mathcal{M}$ is nonparametric. The canonical gradient D_1^Ψ depends on P_0 but this is omitted in the notation because we will only discuss pathwise differentiability at P_0.

A function $f : \mathcal{O}^2 \to \mathbb{R}$ is called (P) one-degenerate if it is symmetric and $Pf(o, \cdot) = 0$. We will use the notation $P^2 f = E_{P^2}[f(O_1, O_2)]$. The parameter Ψ is called second-order pathwise differentiable at P_0 if there exists some symmetric, one-degenerate, P_0^2 square integrable function D_2^Ψ such that

$$\psi_t - \psi_0 = tP_0 D_1^\Psi h_1 + \frac{1}{2}t^2 P_0 D_1^\Psi h_2 + \frac{1}{2}t^2 \int\int D_2^\Psi(o_1, o_2) h_1(o_1) h_1(o_2) dP_0(o_2) dP_0(o_1) + o(t^2).$$

Supplementary Appendix B: Empirical Process Results

We now present two results from empirical process theory, the first of which can be used to control the U-processes that we deal with in the main text when $\mathcal{H}_0$ holds, and the second of which can be used to establish an empirical process condition that is used when $\mathcal{H}_1$ holds.

Before giving an overview of the empirical process theory that we use, we review the notion of a covering number. Let μ be a probability measure over $\mathcal{Z}$. For a class of functions $f : \mathcal{Z} \to \mathbb{R}$ with envelope F (i.e., $|f(z)| \le F(z)$ for all $z \in \mathcal{Z}$), where $0 < \|F\|_{2,\mu} < \infty$, define the covering number $N(\epsilon, \mu, \mathcal{F}, F)$ as the cardinality of the smallest subcollection $\mathcal{F}^* \subseteq \mathcal{F}$ such that, for all $f \in \mathcal{F}$, $\min_{f^* \in \mathcal{F}^*} \|f - f^*\|_{2,\mu} \le \epsilon \|F\|_{2,\mu}$.

Supplementary Appendix B.1: Bounding U-processes

When $\mathcal{H}_0$ holds, our proofs rely on $\mathbb{U}_n(\tilde{\Gamma}_n - \Gamma_0) = o_{P_0}(n^{-1})$ for our method to control the type I error rate. This rate turns out to be plausible, but requires techniques which are different from the now classical empirical process techniques which can be used to establish that $(P_n - P_0)(f_n - f_0) = o_{P_0}(n^{-1/2})$ provided $P_0(f_n - f_0)^2 \to 0$ in probability.

We ignore measurability concerns in this appendix with the understanding that minor modifications are needed to make these results rigorous.

We remind the reader that a function $g : \mathcal{O}^2 \to \mathbb{R}$ is called one-degenerate if and only if g is symmetric in its arguments and $P_0 g(o, \cdot) = 0$ for all $o \in \mathcal{O}$. Let $\mathcal{G}$ denote a collection of one-degenerate functions mapping from $\mathcal{O}^2$ to $\mathbb{R}$, where $\sup_g |g(o_1, o_2)| < G(o_1, o_2)$ for all o_1, o_2 and some envelope function $G \in L^2(P_0)$.

Suppose we wish to estimate some $g_0 \in \mathcal{G}$. We are given a sequence of estimates $\hat{g}_n \in \mathcal{G}$ that is consistent for g_0. Our objective is to show that

$$n\mathbb{U}_n(\hat{g}_n - g_0) = o_{P_0}(1).$$

The uniform entropy integral of $\mathcal{G}$ is given by

$$J(t, \mathcal{G}, G) \triangleq \sup_Q \int_0^t \log N(\epsilon, Q, \mathcal{G}, G) d\epsilon, \tag{A.1}$$

where the supremum is over all distributions Q with support $\mathcal{O}^2$ and $\|G\|_{Q,2} > 0$. We note that the above definition of the entropy integral upper bounds the covering integral given by Nolan and Pollard [1987], which considers a particular choice of Q. The entropy integral above lacks the square root around the logarithm in the integral that is seen in the standard definition of the uniform entropy integral used to bound empirical processes [see, e.g., van der Vaart and Wellner, 1996].

For each $g \in \mathcal{G}$, let H_g represent the Hilbert-Schmidt operator on $L^2(P_0)$ given by $(H_g f)(o) = Pg(o, \cdot) f(\cdot)$. Let $\{W_j : j = 1, 2, ...\}$ be a sequence of i.i.d. standard normal random variables and $\{w_j : j = 1, 2, ...\}$ be an orthonormal basis of $L^2(P_0)$. Let $\tilde{Q}$ be a process on $\mathcal{G}$ defined by

$$\tilde{Q}(g) = \sum_{j=1}^{\infty} \langle H_g w_j, w_j \rangle (W_j^2 - 1) + \sum_{i \neq j} \langle H_g w_j, w_i \rangle W_i W_j.$$

A functional $M : \mathcal{G} \to \mathbb{R}$ is said to belong to $C(\mathcal{G}, P_0^2)$ if $g \mapsto M(g)$ is uniformly continuous for the $L^2(P_0^2)$ seminorm and $\sup_{\mathcal{G}} |M(g)| < \infty$.

We have modified the statement from Nolan and Pollard [1988] slightly to apply to the entropy integral given in (A.1). We omit an analogue to condition (ii) from Nolan and Pollard's statement of the theorem below because it is implied by our strengthening of their condition (i).

THEOREM A.1 (THEOREM 7, NOLAN AND POLLARD, 1988). *Suppose that the one-degenerate class $\mathcal{G}$ satisfies*

(i') $J(1, \mathcal{G}, G) < \infty$;

(iii') $\sup_Q \log N(\epsilon, Q \times P_0, \mathcal{G}, G) < \infty$ *for each* $\epsilon > 0$, *where the supremum is over distributions Q with support $\mathcal{O}$.*

Then there is a version of $\tilde{Q}$ with continuous sample paths in $C(\mathcal{G}, P_0 \times P_0)$ and $n\mathbb{U}_n$ converges in distribution to $\tilde{Q}$.

We will use the following corollary to control the cross-terms.

COROLLARY A.1. *Suppose that $\mathcal{G}$ satisfies the conditions of A.1 and $\hat{g}_n$ is a sequence of one-degenerate random functions that take their values in $\mathcal{G}$ such that $P_0^2(\hat{g}_n - g_0)^2 \to 0$ in probability for some $g_0 \in \mathcal{G}$. Then $n\mathbb{U}_n(\hat{g}_n - g_0) \to 0$ in probability.*

The proof relies on the continuity of sample paths of (a version of) $\tilde{Q}$. The proof is omitted, but we refer the reader to the proof of Lemma 19.24 in van der Vaart [1998] for the analogous empirical process result.

Supplementary Appendix B.2: Controlling $\int \Gamma_n(\cdot, o) dP_0(o)$

We now give sufficient conditions under which $o \mapsto \int \Gamma_n(o_1, o) dP_0(o)$ belongs to a fixed Donsker class with probability approaching one. We recall from van der Vaart and Wellner [1996] that a class $\mathcal{F}$ of functions mapping from $\mathcal{O}$ to $\mathbb{R}$ is Donsker if its uniform entropy integral is finite, which holds if its covering number grows sufficiently slowly as the approximation becomes arbitrarily precise.

Let $\mathcal{G}_2$ be some class of functions $g : \mathcal{O}^2 \to [-M, M]$, $M < \infty$, that contains $\{(o_1, o_2) \mapsto \Gamma_n(o_1, o_2) : \Gamma_n\}$. Without loss of generality, we suppose that $M = 1$. We take the constant function $G_2 \equiv 1$ as envelope for $\mathcal{G}_2$. Let $\mathcal{G}_1 \triangleq \{o_1 \mapsto \int g_2(o_1, o_2) dP_0(o_2) : g_2 \in \mathcal{G}_2\}$, and note that this class similarly has envelope $G_1 \equiv 1$. The main observation of this subappendix is that

$$\sup_{Q_1} N(\epsilon, Q_1, \mathcal{G}_1, G_1) \leq \sup_{Q_2} N(\epsilon, Q_2, \mathcal{G}_2, G_2), \tag{A.2}$$

where the supremum on the left is over all distributions Q_1 on $\mathcal{O}$ such that $\|G_1\|_{2,Q_1} > 0$ and the supremum on the right is over all distributions Q_2 on $\mathcal{O}^2$ such that $\|G\|_{2,Q_2} > 0$. If we can show this, then the uniform entropy integrals are also ordered [Section 2.5 in van der Vaart and Wellner, 1996]:

$$\int_0^\infty \sup_{Q_1} \sqrt{\log N(\epsilon, Q_1, \mathcal{G}_1, G_1)} d\epsilon \leq \int_0^\infty \sup_{Q_2} \sqrt{\log N(\epsilon, Q_2, \mathcal{G}_2, G_2)} d\epsilon, \tag{A.3}$$

where the left- and right-hand sides are the uniform entropy integrals of $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively. Hence, it will suffice to show that the right-hand side is finite. This can be accomplished using the variety of techniques given in Chapter 2 of van der Vaart and Wellner [1996].

We now establish (A.2). Fix a measure Q_1 over $\mathcal{O}$. Let Q_2 represent the product measure $Q_1 \times P_0$. Fix $\epsilon > 0$. Let $g_{2,1}, \ldots, g_{2,m}$ be an $\epsilon \|G_2\|_{2,Q_2}$ cover of $\mathcal{G}_2$ under $\|\cdot\|_{2,Q_2}$ so that $\min_j \|g_2 - g_{2,j}\|_{2,Q_2} < \epsilon \|G_2\|_{2,Q_2}$, where we take m to be equal to its minimal possible value $N(\epsilon, Q_2, \mathcal{G}_2, G_2)$. For each j, let $g_{1,j} \equiv \int g_{2,j}(o_1, o_2) dP_0(o_2)$. Fix $g_1 \in \mathcal{G}_1$. Recall that, by the definition of $\mathcal{G}_1$, there exists a $g_2 \in \mathcal{G}_2$ such that $g_1(\cdot) = \int g_2(\cdot, o) dP_0(o)$. Let j^* be such that $\|g_2 - g_{2,j^*}\|_{2,Q_2} \leq \epsilon \|G_2\|_{2,Q_2}$ for this g_2. Observe that

$$\begin{aligned} \|g_1 - g_{1,j^*}\|_{2,Q_1}^2 &= \int \left(\int [g_2(o_1, o_2) - g_{2,j^*}(o_1, o_2)] \, dP_0(o_2) \right)^2 dQ_1(o_1) \\ &\leq \int [g_2(o_1, o_2) - g_{2,j^*}(o_1, o_2)]^2 \, dQ_2(o_1, o_2) = \|g_2 - g_{2,j^*}\|_{2,Q_2}^2. \end{aligned}$$

By the choice of j^*, it follows that $\|g_1 - g_{1,j^*}\|_{2,Q_1} \leq \epsilon \|G_2\|_{2,Q_2} = \epsilon \|G_1\|_{2,Q_1}$, where we used that $G_1 \equiv 1$ and $G_2 \equiv 1$. That is, $g_{1,1}, \ldots, g_{1,m}$ is an $\epsilon \|G_1\|_{2,Q_1}$ cover of $\mathcal{G}_1$ under $\|\cdot\|_{2,Q_1}$. Thus, $N(\epsilon, Q_1, \mathcal{G}_1, G_1) \leq m$. Recalling that we took $m = N(\epsilon, Q_2, \mathcal{G}_2, G_2)$, we have shown that $N(\epsilon, Q_1, \mathcal{G}_1, G_1) \leq N(\epsilon, Q_2, \mathcal{G}_2, G_2)$. As Q_1 was arbitrary, for each Q_1 with support $\mathcal{O}$ there exists a Q_2 with support $\mathcal{O}^2$ such that the the preceding inequality holds. Hence, (A.2) holds, and thus so too does the uniform entropy integral ordering (A.3).

Supplementary Appendix C: proofs

For any $T \in \mathcal{S}$, we will use the shorthand notation $T_t \triangleq T_{P_t}$, $\frac{d}{dt} T_t \big|_{t=\tilde{t}} \triangleq \dot{T}_{\tilde{t}}$ and $\frac{d^2}{dt^2} T_t \Big|_{t=\tilde{t}} \triangleq \ddot{T}_{\tilde{t}}$. Throughout the appendix we use the following fluctuation submodel through P_0 for pathwise differentiability proofs:

$$\begin{aligned} dP_t(o) &\triangleq \left(1 + t h_1(o) + t^2 h_2(o)\right) dP_0(o), \\ &\text{where } P_0 h_j = 0 \text{ and } \sup_{o \in \mathcal{O}} |h_j(o)| < \infty, j = 1, 2. \end{aligned} \tag{A.4}$$

Proofs for Section 2

We give two lemmas before proving Theorem 1.

LEMMA A.1. *For any $T, U \in \mathcal{S}$ and any fluctuation submodel $dP_t = \left(1 + t h_1 + t^2 h_2\right) dP_0$, we have that, for all $\tilde{t}$ in a neighborhood of zero,*

$$\dot{\Phi}_{\tilde{t}}^{TU} = \int \left[\int e^{-[T_{\tilde{t}}(x_1^T) - U_{\tilde{t}}(x_2^U)]^2} dP_{\tilde{t}}(x_1^T) \right] \left[h_1(o_2) + 2\tilde{t} h_2(o_2) \right] dP_0(o_2)$$

$$
\begin{aligned}
&+\int\left[\int e^{-[T_{\tilde t}(x_1^T)-U_{\tilde t}(x_2^U)]^2}dP_{\tilde t}(x_2^U)\right]\left[h_1(o_1)+2\tilde t h_2(o_1)\right]dP_0(o_1)\\
&-2\iint\left[T_{\tilde t}(x_1^T)-U_{\tilde t}(x_2^U)\right]\left[\frac{d}{dt}T_t(x_1^T)\Big|_{t=\tilde t}-\frac{d}{dt}U_t(x_2^U)\Big|_{t=\tilde t}\right]e^{-[T_{\tilde t}(x_1^T)-U_{\tilde t}(x_2^U)]^2}dP_{\tilde t}(x_2^U)dP_{\tilde t}(x_1^T).
\end{aligned}
$$

PROOF (PROOF OF A.1). We have that

$$
\begin{aligned}
\dot\Phi_{\tilde t}^{TU} &= \frac{d}{dt}\iint e^{-[T_t(x_1^T)-U_t(x_2^U)]^2}\left\{\prod_{j=1}^{2}\left[1+th_1(o_j)+t^2h_2(o_j)\right]\right\}dP_0(o_2)dP_0(o_1)\Bigg|_{t=\tilde t}\\
&=\iint\frac{d}{dt}\,e^{-[T_t(x_1^T)-U_t(x_2^U)]^2}\left\{\prod_{j=1}^{2}\left[1+th_1(o_j)+t^2h_2(o_j)\right]\right\}\Bigg|_{t=\tilde t}dP_0(o_2)dP_0(o_1)\ ,
\end{aligned}
$$

where the derivative is passed under the integral in view of (S2). The result follows by the chain rule.

For each $T,U\in\mathcal{S}$, define

$$
\begin{aligned}
D^{TU}(o)\ \triangleq\ &-2\Phi^{TU}(P_0)+\int\left\{2\left[U_0(o_1)-T_0(o)\right]D_0^T(o)+1\right\}e^{-[T_0(o)-U_0(o_1)]^2}dP_0(o_1)\\
&+\int\left\{2\left[T_0(o_1)-U_0(o)\right]D_0^U(o)+1\right\}e^{-[T_0(o_1)-U_0(o)]^2}dP_0(o_1)\ .
\end{aligned}
$$

We have omitted the dependence of D^{TU} on P_0 in the notation. We first give a key lemma about the parameter Φ^{TU}.

LEMMA A.2 (FIRST-ORDER CANONICAL GRADIENT OF Φ^{TU}). *Let T and U be members of $\mathcal{S}$. Then Φ^{TU} has canonical gradient D^{TU} at P_0.*

PROOF (PROOF OF A.2). To consider first-order behavior it suffices to consider fluctuation submodels in which $h_2(o)=0$ for all o. We first derive the first-order pathwise derivative of the parameter Φ^{TU} at P_0. Applying the preceding lemma at $\tilde t=0$ yields that

$$
\begin{aligned}
\frac{d}{dt}\Phi^{TU}(P_t)\Big|_{t=0} &= \int\left[\int e^{-[T_0(x_1^T)-U_0(x_2^U)]^2}dP_0(x_1^T)\right]h_1(o_2)dP_0(o_2)\\
&+\int\left[\int e^{-[T_0(x_1^T)-U_0(x_2^U)]^2}dP_0(x_2^U)\right]h_1(o_1)dP_0(o_1)\\
&-2\int\int(T_0(x_1^T)-U_0(x_2^U))(\dot T_0(x_1^T)-\dot U_0(x_2^U))e^{-[T_0(x_1^T)-U_0(x_2^U)]^2}dP_0(x_2^U)dP_0(x_1^T).
\end{aligned}
$$

The first two terms in the last equality are equal to

$$
\begin{aligned}
\text{First term} &= \int\left(E_{P_0}\left[e^{-[T_0(X^T)-U_0(x^U)]^2}\right]-E_{P_0^2}\left[e^{-[T_0(X_1^T)-U_0(X_2^U)]^2}\right]\right)h_1(o)dP_0(o)\\
\text{Second term} &= \int\left(E_{P_0}\left[e^{-[T_0(x^T)-U_0(X^U)]^2}\right]-E_{P_0^2}\left[e^{-[T_0(X_1^T)-U_0(X_2^U)]^2}\right]\right)h_1(o)dP_0(o).
\end{aligned}
$$

We now look to find the portion of the canonical gradient given by the third term. We have that

$$
\begin{aligned}
-2\int\int &(T_0(x_1^T) - U_0(x_2^U))\dot{T}_0(x_1^T)e^{-[T_0(x_1^T)-U_0(x_2^U)]^2}dP_0(x_2^U)dP_0(x_1^T) \\
&= \int 2E_{P_0}\left[(U_0(X^U) - T_0(x^T))e^{-[T_0(x^T)-U_0(X^U)]^2}\right]D_0^T(o)h_1(o)dP_0(o) \\
2\int\int &(T_0(x_1^T) - U_0(x_2^U))\dot{U}_0(x_2^U)e^{-[T_0(x_1^T)-U_0(x_2^U)]^2}dP_0(x_2^U)dP_0(x_1^T) \\
&= \int 2E_{P_0}\left[(T_0(X^T) - U_0(x^U))e^{-[T_0(X^T)-U_0(x^U)]^2}\right]D_0^U(o)h_1(o)dP_0(o).
\end{aligned}
$$

Collecting terms, a first-order Taylor expansion of $t \mapsto \Phi^{TU}(P_t)$ about $t = 0$ yields that

$$\Phi^{TU}(P_t) - \Phi^{TU}(P_0) = tE_{P_0}\left[D^{TU}(O)h_1(O)\right] + o(t).$$

Thus Φ^{TU} has canonical gradient D^{TU} at P_0.

The proof of Theorem 1 is simple given the above lemma.

PROOF (PROOF OF THEOREM 1). A.2, the fact that $\Psi(P) \triangleq \Phi^{RR}(P) - 2\Phi^{RS}(P) + \Phi^{SS}(P)$, and the linearity of differentiation immediately yield that the canonical gradient of Ψ can be written as $D^{RR} - 2D^{RS} + D^{SS}$. Straightforward calculations show that this is equivalent to $o \mapsto 2[P_0\Gamma_0(o,\cdot) - \psi_0]$.

We will use the following lemma in the proof of 1 to prove that $R_0(O)$ and $S_0(O)$ are degenerate if $D_1^\Psi \equiv 0$ and $\mathcal{H}_0$ does not hold. Because we were unable to find the proof that the U-statistic kernel for estimating the MMD of two variables X and Y is degenerate if and only if $\mathcal{H}_0$ holds or X and Y are degenerate, we give a proof here that applies in a more general setting than that which we consider in this paper.

LEMMA A.3. *Let Q be a distribution over $(X, Y) \in \mathcal{Z}^2$, where $\mathcal{Z}$ is a compact metric space. Let $(x, y) \mapsto k(x, y)$ be a universal kernel on this metric space, i.e. a kernel for which the resulting reproducing kernel Hilbert space $\mathcal{H}$ is dense in the set of continuous funtions on $\mathcal{Z}$ with respect to the supremum metric. Further, suppose that $E_Q\sqrt{k(X,X)}$ and $E_Q\sqrt{k(Y,Y)}$ are finite. Finally, suppose that the marginal distribution of X under Q is different from the marginal distribution of Y under Q.*

There exists some fixed constant C such that

$$\int \langle\phi(x_1) - \phi(y_1), \phi(x_2) - \phi(y_2)\rangle_{\mathcal{H}} dQ(x_2, y_2) = C \tag{A.5}$$

for (Q almost) all $(x_1, y_1) \in \mathcal{Z}^2$ if and only if the joint distribution of (X, Y) under Q is degenerate at a single point. Above $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and $\phi(z) \triangleq k(z,\cdot)$ are the inner product and the feature map in $\mathcal{H}$, respectively.

PROOF. If Q is degenerate then clearly (A.5) holds.

If (A.5) holds, then our assumption that X has a different marginal distribution than Y tells us that $C > 0$ [Gretton et al., 2012]. Hence, for almost all (x_1, y_1),

$$\langle\phi(x_1) - \phi(y_1), \mu_X - \mu_Y\rangle_{\mathcal{H}} - \langle\mu_X - \mu_Y, \mu_X - \mu_Y\rangle_{\mathcal{H}} = 0,$$

where μ_X and μ_Y in $\mathcal{H}$ have the property that $\langle \mu_X, f\rangle_{\mathcal{H}} = E_Q f(X)$ and $\langle \mu_Y, f\rangle_{\mathcal{H}} = E_Q f(Y)$ for all $f \in \mathcal{H}$ [Lemma 3 in Gretton et al., 2012]. The above holds if and only if $\phi(x_1) - \phi(y_1) = \mu_X - \mu_Y$. Noting that $\mu_X - \mu_Y$ does not rely on x_1, y_1, it follows that $\phi(x_1) - \phi(y_1)$ must not rely on x_1, y_1 for all (x_1, y_1) in some Q probability one set $\mathcal{D} \subseteq \mathcal{Z}^2$.

Fix a continuous function $f : \mathcal{Z} \to \mathbb{R}$ and $x_1, y_1 \in \mathcal{D}$. For any $\epsilon > 0$, the universality of $\mathcal{H}$ ensures that there exists an $f_\epsilon \in \mathcal{H}$ such that $\|f_\epsilon - f\|_\infty \leq \epsilon$. By the triangle inequality,

$$|f(x_1) - f(y_1) - f_\epsilon(x_1) + f_\epsilon(y_1)| \leq 2\epsilon.$$

Because $\phi(x_1) - \phi(y_1)$ is constant and $f \in \mathcal{H}$, $\langle \phi(x_1) - \phi(y_1), f_\epsilon\rangle_{\mathcal{H}} = f_\epsilon(x_1) - f_\epsilon(y_1)$ does not rely on x_1, y_1 for any ϵ. Furthermore, the fact that f_ϵ converges to f in supremum norm ensures that $|f_\epsilon(x_1) - f_\epsilon(y_1)|$ converges to a fixed quantity K (which does not rely on x_1 or y_1) as $\epsilon \to 0$. Applying this to the above yields that $f(x_1) - f(y_1) = K$.

As f was an arbitrary continuous function and $X_1 \not\equiv Y_1$, we can apply this relation to $z \mapsto z$ and $z \mapsto z^2$ to show that $x_1 - y_1$ and $x_1 + y_1$ do not rely on the choice of $(x_1, y_1) \in \mathcal{D}$. Hence $(x_1 - y_1 + x_1 + y_1)/2 = x_1$ and $(x_1 + y_1 - y_1 + x_1)/2 = y_1$ do not rely on the choice of $(x_1, y_1) \in \mathcal{D}$. This can only occur if (x_1, y_1) are constant over the probability 1 set $\mathcal{D}$, i.e. if Q is degenerate.

For the two-sample problem in Gretton et al. [2012], one can take Q to be a product distribution of the marginal distribution of X and the marginal distribution of Y.

PROOF (PROOF OF 1). We first prove sufficiency. If (i) holds, then $2D^{RS} = D^{RR} + D^{SS}$. It follows that $D_1^\Psi \equiv 0$ under $\mathcal{H}_0$. Now suppose (ii) holds. It is a simple matter of algebra to verify that $D_1^{RR} \equiv D_1^{RS} \equiv D_1^{SS} \equiv 0$. Hence $D_1^\Psi \equiv 0$, yielding the sufficiency of the stated conditions.

We now show the necessity of the stated conditions. Suppose that $\sigma_0 = 0$ and $\mathcal{H}_0$ does not hold. It is easy to verify that

$$\begin{aligned}\tilde{D}_1^\Psi \triangleq\ & E_{P_0}\left[e^{-[R_0(O)-R_0(o)]^2}\right] + E_{P_0}\left[e^{-[S_0(O)-S_0(o)]^2}\right] \\ & - E_{P_0}\left[e^{-[R_0(O)-S_0(o)]^2}\right] - E_{P_0}\left[e^{-[R_0(o)-S_0(O)]^2}\right] - \psi_0\end{aligned}$$

is a first-order gradient in the model where R_0 and S_0 are known (possibly an inefficient gradient depending on the form of R and S). Call the variance of this gradient $\tilde{\sigma}_0$. As the model where R_0 and S_0 are known is a submodel of the (locally) nonparametric model, $\tilde{\sigma}_0 \leq \sigma_0$, and hence $\tilde{\sigma}_0 = 0$ and $\tilde{D}_1^\Psi \equiv 0$. Now, if $\tilde{\sigma}_0 = 0$ and $\mathcal{H}_0$ does not hold, then A.3 shows that $R_0(O)$ and $S_0(O)$ are degenerate. Finally, $\tilde{D}_1^\Psi \equiv 0$ and the degeneracy of $R_0(O)$ and $S_0(O)$ shows that for almost all o,

$$D_1^\Psi(o) = 2D^{RS}(o) = 2(s_0 - r_0)\left(D_0^R(o) - D_0^S(o)\right) e^{-[r_0 - s_0]^2},$$

where we use r_0 and s_0 to denote the (probability 1) values of $R_0(O)$ and $S_0(O)$. The above is zero almost surely if and only if $D_0^R \equiv D_0^S$. Thus $\sigma_0 = 0$ only if (i) or (ii) holds.

We give the following lemma before proving Theorem 2. Before giving the lemma, we define the function $\Pi : \mathcal{S} \to \mathbb{R}$. Suppressing the dependence on P_0 and h_1, h_2, for all $V \in \mathcal{S}$ and $t \neq 0$ we define

$$\Pi(V) \triangleq 2\int\int \Big[2(V_0(o_2) - V_0(o_1))\dot{V}_0(o_2)h_1(o_2) + 2(V_0(o_2) - V_0(o_1))^2\dot{V}_0(o_2)^2$$

$$+ h_2(o_2) - \dot{V}_0(o_2)^2 + (V_0(o_2) - V_0(o_1))\ddot{V}_0(o_2)\Big] e^{-[V_0(o_2)-V_0(o_1)]^2} dP_0(o_2)dP_0(o_1).$$

LEMMA A.4. *For any fluctuation submodel consistent with (A.4), $T, U \in \mathcal{S}$ with $T_0(O) \stackrel{d}{=} U_0(O)$, and $t \in \mathbb{R}$ sufficiently close to zero, we have that*

$$\frac{d^2}{dt^2}\Phi^{TU}(P_t)\Big|_{t=0} = 2\int\int \Gamma_0^{TU}(o_1,o_2)h_1(o_1)h_1(o_2)dP_0(o_2)dP_0(o_1) + \Pi(T) + \Pi(U).$$

PROOF. Let $H_t(o) \triangleq 1 + th_1(o) + t^2 h_2(o)$ and $\dot{H}_t(o) \triangleq h_1(o) + 2th_2(o)$.

$$\begin{aligned}\frac{d^2}{dt^2}\Phi^{TU}(P_t)\Big|_{t=0} = \frac{d}{dt}\int\int \Big[&H_t(o_1)\dot{H}_t(o_2) + \dot{H}_t(o_1)H_t(o_2) \\ &- 2(T_t(o_1) - U_t(o_2))\left(\dot{T}_t(o_1) - \dot{U}_t(o_2)\right) H_t(o_1)H_t(o_2)\Big] \\ &\times e^{-[T_t(o_1)-U_t(o_2)]^2} dP_0(o_2)dP_0(o_1)\Big|_{t=0} \qquad \text{(A.6)}\end{aligned}$$

We will pass the derivative inside the integral using (S2) and apply the product rule. The first term we need to consider is

$$\begin{aligned}&\frac{d}{dt}\Big[H_t(o_1)\dot{H}_t(o_2) + \dot{H}_t(o_1)H_t(o_2) - 2(T_t(o_1) - U_t(o_2))\left(\dot{T}_t(o_1) - \dot{U}_t(o_2)\right) H_t(o_1)H_t(o_2)\Big]\Big|_{t=0} \\ &= 2\left[h_2(o_1) + h_1(o_1)h_1(o_2) + h_2(o_2)\right] - 2\left(\dot{T}_0(o_1) - \dot{U}_0(o_2)\right)^2 - 2(T_0(o_1) - U_0(o_2))\left(\ddot{T}_0(o_1) - \ddot{U}_0(o_2)\right) \\ &\qquad - 2(T_0(o_1) - U_0(o_2))\left(\dot{T}_0(o_1) - \dot{U}_0(o_2)\right)(h_1(o_1) + h_1(o_2)).\end{aligned}$$

The second is

$$\frac{d}{dt}e^{-[T_t(o_1)-U_t(o_2)]^2}\Big|_{t=0} = -2(T_0(o_1) - U_0(o_2))\left(\dot{T}_0(o_1) - \dot{U}_0(o_2)\right) e^{-[T_0(o_1)-U_0(o_2)]^2}.$$

Returning to (A.6), this shows that $\frac{d^2}{dt^2}\Phi^{TU}(P_t)\Big|_{t=0}$ is equal to

$$\begin{aligned}&2\int\int\Big[-2(T_0(o_1) - U_0(o_2))\dot{T}_0(o_1)h_1(o_1) + 2(T_0(o_1) - U_0(o_2))^2\dot{T}_0(o_1)^2 \\ &\qquad + h_2(o_1) - \dot{T}_0(o_1)^2 - (T_0(o_1) - U_0(o_2))\ddot{T}_0(o_1)\Big] e^{-[T_0(o_1)-U_0(o_2)]^2} dP_0(o_2)dP_0(o_1) \\ &+ 2\int\int\Big[2(T_0(o_1) - U_0(o_2))\dot{U}_0(o_2)h_1(o_2) + 2(T_0(o_1) - U_0(o_2))^2\dot{U}_0(o_2)^2 \\ &\qquad + h_2(o_2) - \dot{U}_0(o_2)^2 + (T_0(o_1) - U_0(o_2))\ddot{U}_0(o_2)\Big] e^{-[T_0(o_1)-U_0(o_2)]^2} dP_0(o_2)dP_0(o_1) \\ &+ 2\int\int\Big[2(T_0(o_1) - U_0(o_2))\left(\dot{U}_0(o_2)h_1(o_1) - \dot{T}_0(o_1)h_1(o_2)\right) \\ &\qquad - \left(4(T_0(o_1) - U_0(o_2))^2 - 2\right)\dot{T}_0(o_1)\dot{U}_0(o_2) + h_1(o_1)h_1(o_2)\Big] e^{-[T_0(o_1)-U_0(o_2)]^2} dP_0(o_2)dP_0(o_1).\end{aligned}$$

The expression inside the second pair of integrals only depends on o_1 through $T(o_1)$. Thus we can rewrite this term as $E_{P_0}[f(T(O_1))]$ for a fixed function f that relies on P_0, h_1, h_2, and U.

Under $\mathcal{H}_0$, we can rewrite this term as $E_{P_0}[f(U(O_1))]$. That is, we can replace each $T(O_1)$ in the second pair of integrals with $U(O_1)$. This yields $\Pi(U)$. Switching the roles of o_1 and o_2 in the first pair of integrals above and applying Fubini's theorem shows that

$$2\int\int\Big[2(T_0(o_2)-U_0(o_1))\dot{T}_0(o_2)h_1(o_2)+2(T_0(o_2)-U_0(o_1))^2\dot{T}_0(o_2)^2 \\ +h_2(o_2)-\dot{T}_0(o_2)^2+(T_0(o_2)-U_0(o_1))\ddot{T}_0(o_2)\Big]e^{-[T_0(o_2)-U_0(o_1)]^2}dP_0(o_2)dP_0(o_1).$$

By the same arguments used to for the second pair of integrals, the above expression is equal to $\Pi(T)$ under $\mathcal{H}_0$. By (S3), the third pair of integrals can be rewritten as

$$2\int\int\Big[2(T_0(o_1)-U_0(o_2))\left(D_0^U(o_2)-D_0^T(o_1)\right)-\left(4(T_0(o_1)-U_0(o_2))^2-2\right)D_0^T(o_1)D_0^U(o_2)+1\Big] \\ \times e^{-[T_0(o_1)-U_0(o_2)]^2}h_1(o_1)h_1(o_2)dP_0(o_2)dP_0(o_1).$$

PROOF (PROOF OF THEOREM 2). We start by noting that $\frac{1}{2}\frac{d^2}{dt^2}\psi_t\Big|_{t=0}$ is equal to

$$\frac{1}{2}\left[\frac{d^2}{dt^2}\Phi^{TT}(P_t)\Big|_{t=0}+\frac{d^2}{dt^2}\Phi^{UU}(P_t)\Big|_{t=0}-\frac{d^2}{dt^2}\Phi^{TU}(P_t)\Big|_{t=0}-\frac{d^2}{dt^2}\Phi^{UT}(P_t)\Big|_{t=0}\right]$$
$$=\int\int\left[\Gamma_0^{RR}(o_1,o_2)+\Gamma_0^{SS}(o_1,o_2)-\Gamma_0^{RS}(o_1,o_2)-\Gamma_0^{SR}(o_1,o_2)\right]h_1(o_1)h_1(o_2)dP_0(o_2)dP_0(o_1)$$
$$=\frac{1}{2}\int\int D_2^\Psi(o_1,o_2)h_1(o_1)h_1(o_2)dP_0(o_2)dP_0(o_1),$$

where the penultimate equality makes use of A.4. It is easy to verify that $D_2^\Psi(o_1,o_2)=D_2^\Psi(o_2,o_1)$ for all o_1,o_2. The arguments given below the theorem statement in the main text establish the one-degeneracy of Γ_0 under $\mathcal{H}_0$ show that $E_{P_0}[D_2^\Psi(O,o)]=E_{P_0}[D_2^\Psi(o,O)]=0$ for all $o\in\mathcal{O}$ under $\mathcal{H}_0$. Condition (S2) ensures that $\|D_2^\Psi\|_{2,P_0^2}<\infty$, and thus D_2^Ψ is P_0^2 square integrable and one-degenerate.

Because the first pathwise derivative is zero under the null, we have that

$$\psi_t-\psi_0=\frac{1}{2}t^2\int\int D_2^\Psi(o_1,o_2)h(o_1)h(o_2)dP_0(o_1)dP_0(o_2)+o(t^2).$$

Thus D_2^Ψ is a second-order canonical gradient of Ψ at P_0.

We give a lemma before proving Theorem 3.

LEMMA A.5. *Fix $P\in\mathcal{M}$. For all $T,U\in\mathcal{S}$, let*

$$\mathrm{Rem}_P^{\Phi^{TU}}\triangleq\|L_P^{TU}\|_{2,P_0}\|M_P^{TU}\|_{2,P_0}+\|\mathrm{Rem}_P^T\|_{1,P_0}\|\mathrm{Rem}_P^U\|_{1,P_0}+\|M_P^{TU}\|_{4,P_0}^4.$$

There exists a mapping $\zeta(P,P_0,\cdot):\mathcal{S}\to\mathbb{R}$ such that, for all $T,U\in\mathcal{S}$ for which $T_0(O)\overset{d}{=}U_0(O)$,

$$\left|P_0^2\Gamma_P^{TU}-\Phi^{TU}(P_0)-\zeta(P,P_0,T)-\zeta(P,P_0,U)\right|\lesssim\mathrm{Rem}_P^{\Phi^{TU}}$$

PROOF (PROOF OF A.5). In this proof we use $F(P, P_0, T, U)$ to denote any constant which can be written as $\tilde{\zeta}(P, P_0, T) + \tilde{\zeta}(P, P_0, U)$ for expressions $\tilde{\zeta}(P, P_0, T)$ and $\tilde{\zeta}(P, P_0, U)$ which satisfy $\tilde{\zeta}(P, P_0, T) = \tilde{\zeta}(P, P_0, U)$ whenever $T = U$. We will write $c_1 F(P, P_0, T, U) + c_2 F(P, P_0, T, U) = F(P, P_0, T, U)$ for any real numbers c_1, c_2. We then fix ζ to be the final instance of $\tilde{\zeta}$ upon exiting the proof.

Fix $T, U \in \mathcal{S}$. Let $b_0(o_1, o_2) \triangleq T_0(o_1) - U_0(o_2)$ and $b(o_1, o_2) \triangleq T_P(o_1) - U_P(o_2)$ for any o_1, o_2. For ease of notation, in the expected values below we will write B and B_0 to refer to $b(O_1, O_2)$ and $b_0(O_1, O_2)$, respectively. We also write T for $T_P(O_1)$, T_0 for $T_0(O_1)$, Rem_P^T for $\mathrm{Rem}_P^T(O_1)$, U for $U_P(O_2)$, U_0 for $U_0(O_2)$, and Rem_P^U for $\mathrm{Rem}_P^U(O_2)$.

We have that

$$\begin{aligned}
P_0^2\Gamma_P^{TU} - \Phi^{TU}(P_0) &= E_{P_0^2}\left[e^{-B^2} - e^{-B_0^2}\right] + E_{P_0^2}\left[2B\left(D_P^U(O_2) - D_P^T(O_1)\right)e^{-B^2}\right] \\
&\quad - E_{P_0^2}\left[\left(4B^2-2\right)D_P^T(O_1)D_P^U(O_2)e^{-B^2}\right] \\
&= E_{P_0^2}\left[e^{-B^2} - e^{-B_0^2}\right] - E_{P_0^2}\left[2B\left(B_0 - B\right)e^{-B^2}\right] \\
&\quad + E_{P_0^2}\left[2B\left(\mathrm{Rem}_P^U - \mathrm{Rem}_P^T\right)e^{-B^2}\right] - E_{P_0^2}\left[\left(4B^2-2\right)[T-T_0]\,[U-U_0]\,e^{-B^2}\right] \\
&\quad - E_{P_0^2}\left[\left(4B^2-2\right)\left([T-T_0]\,\mathrm{Rem}_P^U + \mathrm{Rem}_P^T\,[U-U_0]\right)e^{-B^2}\right] \\
&\quad - E_{P_0^2}\left[\left(4B^2-2\right)\mathrm{Rem}_P^T\,\mathrm{Rem}_P^U\,e^{-B^2}\right].
\end{aligned}$$

A third-order Taylor expansion of $b_0 \mapsto \exp(-b_0^2)$ about $b_0 = b$ yields

$$e^{-b^2} - e^{-b_0^2} = 2b(b_0-b)e^{-b^2} - \left(2b^2-1\right)(b_0-b)^2e^{-b^2} + \frac{2}{3}b\left(2b^2-3\right)(b_0-b)^3e^{-b^2} + O\left((b_0-b)^4\right),$$

where the magnitude of the $O((b_0-b)^4)$ term is uniformly bounded above by $C(b_0-b)^4$ for some constant $C > 0$ when b_0 and b fall in $[-1, 1]$. For the second-order term, we have

$$\begin{aligned}
E_{P_0^2}\left[-\left(2B^2-1\right)(B_0-B)^2e^{-B^2}\right] &= E_{P_0^2}\left[\left(4B^2-2\right)(T-T_0)\,(U-U_0)\,e^{-B^2}\right] \\
&\quad - E_{P_0^2}\left[\left([T-T_0]^2 + [U-U_0]^2\right)\left(2B^2-1\right)e^{-B^2}\right].
\end{aligned}$$

Thus we have that

$$\begin{aligned}
P_0^2\Gamma_P^{TU} - \Phi^{TU}(P_0) &= E_{P_0^2}\left[2B\left(\mathrm{Rem}_P^U - \mathrm{Rem}_P^T\right)e^{-B^2}\right] \\
&\quad + O\left(\|B - B_0\|_{4,P_0}^4\right) - E_{P_0^2}\left[\left(4B^2-2\right)\mathrm{Rem}_P^T\,\mathrm{Rem}_P^U\,e^{-B^2}\right] \\
&\quad - E_{P_0^2}\left[\left([T-T_0]^2 + [U-U_0]^2\right)\left(2B^2-1\right)e^{-B^2}\right] \\
&\quad + \frac{2}{3}E_{P_0}\left[B\left(2B^2-3\right)(B_0-B)^3e^{-B^2}\right]. \qquad \text{(A.7)}
\end{aligned}$$

A Taylor expansion of $f_1(z) = 2ze^{-z^2}$ shows that there exists a $\tilde{B}_1(o_1, o_2)$ that falls between $B(o_1, o_2)$ and $B_0(o_1, o_2)$ for all o_1, o_2 such that

$$E_{P_0^2}\left[2B\left(\mathrm{Rem}_P^U - \mathrm{Rem}_P^T\right)e^{-B^2}\right] = E_{P_0^2}\left[\left(\mathrm{Rem}_P^U - \mathrm{Rem}_P^T\right)\left(2B_0e^{-B_0^2} + (B-B_0)\dot{f}_1(\tilde{B})\right)\right]$$

$$= F(P, P_0, T, U) + E_{P_0^2}\left[\left(\mathrm{Rem}_P^U - \mathrm{Rem}_P^T\right)(B - B_0)\dot{f}_1(\tilde{B})\right], \tag{A.8}$$

where the second equality holds under $\mathcal{H}_0$. The boundedness of $\dot{f}_1$ in $[-2, 2]$, the triangle inequality, and the Cauchy-Schwarz inequality yield

$$\begin{aligned} E_{P_0^2}\left|\left(\mathrm{Rem}_P^U - \mathrm{Rem}_P^T\right)(B - B_0)\dot{f}_1(\tilde{B})\right| &\lesssim E_{P_0^2}\left|\left(\mathrm{Rem}_P^U - \mathrm{Rem}_P^T\right)(B - B_0)\right| \\ \lesssim E_{P_0^2}\left|L_P^{TU}(O_1)M_P^{TU}(O_2)\right| + E_{P_0}\left|L_P^{TU}\right| E_{P_0}\left|M_P^{TU}\right| &\lesssim \|L_P^{TU}\|_{2,P_0}\|M_P^{TU}\|_{2,P_0}. \end{aligned} \tag{A.9}$$

A Taylor expansion of $f_2(z) = (2z^2 - 1)e^{-z^2}$ yields that there exists a $\tilde{B}_2$ that falls between B and B_0 such that

$$\begin{aligned} &E_{P_0^2}\left[\left([T - T_0]^2 + [U - U_0]^2\right)\left(2B^2 - 1\right)e^{-B^2}\right] \\ &\quad = E_{P_0^2}\left[\left([T - T_0]^2 + [U - U_0]^2\right)\left(2B_0^2 - 1\right)e^{-B_0^2}\right] \\ &\qquad + 2E_{P_0^2}\left[\left([T - T_0]^2 + [U - U_0]^2\right)(B - B_0)\left(B(2B^2 - 3)\right)e^{-B^2}\right] \\ &\qquad + E_{P_0^2}\left[\left([T - T_0]^2 + [U - U_0]^2\right)(B - B_0)^2\frac{\ddot{f}_2(\tilde{B}_2)}{2}\right]. \end{aligned}$$

The first line on the right is equal to $F(P, P_0, T, U)$ under $\mathcal{H}_0$. By the triangle inequality and the boundedness of $\ddot{f}_2$ on $[-2, 2]$, the third line satisfies

$$\begin{aligned} E_{P_0^2}\left[\left([T - T_0]^2 + [U - U_0]^2\right)(B - B_0)^2\frac{\ddot{f}_2(\tilde{B}_2)}{2}\right] &\lesssim \sum_{k=0}^{4} E_{P_0^2}\left|[T - T_0]^k [U - U_0]^{4-k}\right| \\ \lesssim \sum_{k=0}^{4} E_{P_0}\left|[M_P^{TU}]^k\right| E_{P_0}\left|[M_P^{TU}]^{4-k}\right| &\lesssim \|M_P^{TU}\|_{4,P_0}^4. \end{aligned} \tag{A.10}$$

The final inequality above holds by the FKG inequality [Fortuin et al., 1971]. It follows that

$$\begin{aligned} &E_{P_0^2}\left[\left([T - T_0]^2 + [U - U_0]^2\right)\left(2B^2 - 1\right)e^{-B^2}\right] + \frac{2}{3}E_{P_0}\left[B\left(2B^2 - 3\right)(B_0 - B)^3 e^{-B^2}\right] \\ &\quad = \frac{4}{3}E_{P_0^2}\left[\left([T - T_0]^3 - [U - U_0]^3\right)B(2B^2 - 3)e^{-B^2}\right] + F(P, P_0, T, U) + O(\|M_P^{TU}\|_{4,P_0}^4) \\ &\quad = F(P, P_0, T, U) + O(\|M_P^{TU}\|_{4,P_0}^4), \end{aligned} \tag{A.11}$$

where the final equality holds under $\mathcal{H}_0$ by a Taylor expansion of $z \mapsto z(2z^2 - 3)e^{-z^2}$ and analogous calculations to those used in (A.10). We note that the second equality above uses a different F and a different big-O term than the line above, and that the big-O term can be upper bounded by $C\|M_P^{TU}\|_{4,P_0}^4$ for a constant $C > 0$.

Plugging (A.8), (A.9), and (A.11) into (A.7), applying the triangle inequality, and using the bounds on B gives the result.

We give a lemma before proving Theorem 3.

LEMMA A.6. *Let $K_P \triangleq \|L_P^{RS}\|_{1,P_0} + \|M_P^{RS}\|_{2,P_0}^2$ for all $P \in \mathcal{M}$. If $\mathcal{H}_0$ holds, then for all $P \in \mathcal{M}$,*

$$\sup_{o_1 \in \mathcal{O}'} |P_0 \Gamma_P(o_1, \cdot)| \lesssim K_P,$$

where $\mathcal{O}' \subseteq \mathcal{O}$ is some P_0 probability 1 set. More generally, for all $P_0 \in \mathcal{M}$,

$$\left|P_0^2 \Gamma_P - \psi_0\right| \lesssim K_P.$$

PROOF (PROOF OF A.6). For $T, U \in \mathcal{S}$, we have that

$$\Gamma_P^{TU} = \left[1 + 2(T_P - U_P) D_P^U\right] e^{-[T_P - U_P]^2} - 2\left[(T_P - U_P) + \left(2(T_P - U_P)^2 - 1\right) D_P^U\right] D_P^T e^{-[T_P - U_P]^2}.$$

Above we have omitted the dependence of Γ^{TU} on (o_1, o_2), T and D_P^T on o_1, and U and D_P^U on o_2. For P_0 almost all $o_1 \in \mathcal{O}$, $P_0 \Gamma_P^{TU}(o_1, \cdot)$ is equal to

$$\begin{aligned}
&P_0\left[1 + 2(T_P(o_1) - U_P)(U_P - U_0)\right] e^{-[T_P(o_1) - U_P]^2} + O\left(\|\mathrm{Rem}_P^U\|_{1,P_0}\right) \\
&\quad - 2P_0\left[(T_P(o_1) - U_P) + \left(2(T_P(o_1) - U_P)^2 - 1\right)(U_P - U_0)\right] D_P^T(o_1) e^{-[T_P(o_1) - U_P]^2}
\end{aligned}$$

where the magnitude of the big-O remainder term is upper bounded by $C\|\mathrm{Rem}_P^U\|_{1,P_0}$ for a constant $C > 0$ which does not depend on o_1. Taylor expansions of the first and third terms above yield

$$\begin{aligned}
P_0 \Gamma_P^{TU}(o_1, \cdot) = &P_0 e^{-[T_P(o_1) - U_0]^2} - 2P_0(T_P(o_1) - U_0) D_P^T(o_1) e^{-[T_P(o_1) - U_0]^2} \\
&+ O\left(\|\mathrm{Rem}_P^U\|_{1,P_0}\right) + O\left(\|U_P - U_0\|_{2,P_0}^2\right),
\end{aligned}$$

where the magnitude of the big-O term can be upper bounded by $C\|U_P - U_0\|_{2,P_0}^2$. If $T_0(O) \stackrel{d}{=} U_0(O)$, then

$$\begin{aligned}
P_0 \Gamma_P^{TU}(o_1, \cdot) = &P_0 e^{-[T_P(o_1) - T_0]^2} - 2P_0(T_P(o_1) - T_0) D_P^T(o_1) e^{-[T_P(o_1) - T_0]^2} \\
&+ O\left(\|\mathrm{Rem}_P^U\|_{1,P_0}\right) + O\left(\|U_P - U_0\|_{2,P_0}^2\right).
\end{aligned}$$

Recall that $T, U \in \mathcal{S}$ were arbitrary. Using that $\Gamma_P \triangleq \Gamma_P^{RR} - \Gamma_P^{RS} - \Gamma_P^{SR} + \Gamma_P^{SS}$ and applying the triangle inequality gives the first result.

We now turn to the second result. For any $T, U \in \mathcal{S}$ and $P \in \mathcal{M}$, we have that

$$\begin{aligned}
P_0^2 \Gamma_P^{TU} = &\Big[2(T_P - U_P)(U_0 - U_P - T_0 + T_P) + 1 \\
&\quad - \left(4(T_P - U_P)^2 - 2\right)(U_P - U_0)(T_P - T_0)\Big] e^{-[T_P - U_P]^2} + O\left(\|L_P^{TU}\|_{1,P_0}\right) \\
= &\left[2(T_P - U_P)(U_0 - U_P - T_0 + T_P) + 1\right] e^{-[T_P - U_P]^2} + O\left(\|L_P^{TU}\|_{1,P_0}\right) + O\left(\|M_P^{TU}\|_{2,P_0}^2\right) \\
= &\Phi^{TU}(P_0) + O\left(\|L_P^{TU}\|_{1,P_0}\right) + O\left(\|M_P^{TU}\|_{2,P_0}^2\right),
\end{aligned}$$

where the final equality holds by a first-order Taylor expansion of $(t, u) \mapsto e^{-[t-u]^2}$. The fact that $\Gamma_P \triangleq \Gamma_P^{RR} - 2\Gamma_P^{RS} + \Gamma_P^{SS}$ yields the result.

PROOF (PROOF OF THEOREM 3). Fix $P \in \mathcal{M}$ and let P_0 satisfy $\mathcal{H}_0$. We have that

$$P_0^2\Gamma_P - \psi_0 = P_0^2\Gamma_P^{RR} - \Phi^{RR}(P_0) + P_0^2\Gamma_P^{SS} - \Phi^{SS}(P_0) - \left[P_0^2\Gamma_P^{RS} - \Phi^{RS}(P_0) + P_0^2\Gamma_P^{SR} - \Phi^{SR}(P_0)\right].$$

Taking the absolute value of both sides, applying the triangle inequality, and using A.5 yields

$$\left|P_0^2\Gamma_P - \psi_0\right| \lesssim \mathrm{Rem}_P^{\Phi^{RR}} + \mathrm{Rem}_P^{\Phi^{SS}} + 2\,\mathrm{Rem}_P^{\Phi^{RS}} \lesssim \|L_P^{RS}\|_{1,P_0}^2 + \|M_P^{RS}\|_{4,P_0}^4 + \|L_P^{RS}\|_{2,P_0}\|M_P^{RS}\|_{2,P_0},$$

where the final inequality uses the maximum in the definition of L_P^{RS} and M_P^{RS}.

The inequality for when P_0 satisfies $\mathcal{H}_1$ is proven in A.6.

0.1. Proofs for Section 3

PROOF (PROOF OF 1). By the first result of A.6, $|P_0\Gamma_n(o_1,\cdot)| \lesssim K_n$ for P_0 almost all $o_1 \in \mathcal{O}'$. We have that

$$|(P_n - P_0)P_0\Gamma_n| = K_n\left|(P_n - P_0)\left(\frac{P_0\Gamma_n}{K_n}\right)\right|.$$

The fact that $\left\{o_1 \mapsto \frac{P_0\Gamma_n(o_1,\cdot)}{K_n} : \hat{P}_n\right\}$ belongs to a P_0 Donsker class with probability approaching 1, where $\hat{P}_n$ varies over the set of its possible realizations, yields that $(P_n - P_0)\left(\frac{P_0\Gamma_n}{K_n}\right) = O_{P_0}(n^{-1/2})$ [van der Vaart and Wellner, 1996], and thus the right-hand side above is $O_{P_0}(K_n/\sqrt{n})$. If $K_n = o_{P_0}(n^{-1/2})$, then this yields that the right-hand side above is $o_{P_0}(n^{-1})$.

PROOF (PROOF OF THEOREM 4). Plugging C1), C2), and C3) into (5) yields

$$\psi_n - \psi_0 = \mathbb{U}_n\Gamma_0 + o_{P_0}(n^{-1}). \tag{A.12}$$

By Section 5.5.2 of Serfling [1980] and the fact that Γ_0 is P_0 degenerate and uniformly bounded, $n\mathbb{U}_n\Gamma_0 \rightsquigarrow \sum_{k=1}^{\infty}\lambda_k(Z_k^2 - 1)$.

We now prove that all of the eigenvalues of $h(o) \mapsto E_{P_0}\left[\tilde{\Gamma}_0(O,o)h(O)\right]$ are nonnegative. Consider a submodel $\{P_t : t\}$ with first-order score $h_1 \in L^2(P_0)$ and second-order score $h_2 \equiv 0$. By the second-order pathwise differentiability of Ψ,

$$\frac{\psi_t - \psi_0}{t^2} = \frac{1}{2}\int\int D_2^{\Psi}(o_1,o_2)h_1(o_1)h_1(o_2)dP_0(o_1)dP_0(o_2) + o(1).$$

The left-hand side is nonnegative for all $t \neq 0$ since $\psi_t \geq 0 = \psi_0$ under $\mathcal{H}_0$. Thus taking the limit inferior as $t \to 0$ of both sides shows that

$$\frac{1}{2}\int\int D_2^{\Psi}(o_1,o_2)h_1(o_1)h_1(o_2)dP_0(o_1)dP_0(o_2) \geq 0.$$

Using that $\tilde{\Gamma}_0 = \Gamma_0$ under $\mathcal{H}_0$ and $\Gamma_0 = \frac{1}{2}D_2^{\Psi}$, we have that $\langle o \mapsto E_{P_0}[\tilde{\Gamma}_0(O,o)h_1(O)], h_1\rangle \geq 0$, where the inner product is that of $L^2(P_0)$. For any $h_1 \in L^2(P_0)$, it is well known that one can choose a submodel P_t with first-order score $h_1 \in L^2(P_0)$. Hence the above relation holds for all $h_1 \in L^2(P_0)$ and all of the eigenvalues of $h(o) \mapsto E_{P_0}\left[\tilde{\Gamma}_0(O,o)h(O)\right]$ are nonnegative.

PROOF (PROOF OF 2). In this case $\Gamma_0(o_1, o_2) = 2D_0^R(o_1)D_0^R(o_2)$ under $\mathcal{H}_0$. The central limit theorem yields that $\sigma_1^{-1}\sqrt{n}(P_n - P_0)D_0^R \rightsquigarrow Z$. By the continuous mapping theorem, $\sigma_1^{-2}n(P_n - P_0)^2\Gamma_0/2 \rightsquigarrow Z^2$. Now use that

$$\begin{aligned}\frac{n\mathbb{U}_n\Gamma_0}{2\sigma_1^2} &= \frac{n}{2\sigma_1^2(n-1)}\left[n(P_n - P_0)^2\Gamma_0 - \frac{1}{n}\sum_{i=1}^n \Gamma_0(O_i, O_i)\right] \\ &= \frac{n}{2\sigma_1^2(n-1)}\left[n(P_n - P_0)^2\Gamma_0 - \frac{2}{n}\sum_{i=1}^n D_0^R(O_i)^2\right].\end{aligned}$$

The above quantity converges in distribution to $Z^2 - 1$ by the weak law of large numbers and Slutsky's theorem.

PROOF (PROOF OF THEOREM 6). We have

$$\begin{aligned}\psi_n &= 2(P_n - P_0)P_0\Gamma_n + P_0^2\Gamma_n + \mathbb{U}_n\tilde{\Gamma}_n \\ &= 2(P_n - P_0)P_0\Gamma_0 + P_0^2\Gamma_n + \mathbb{U}_n\tilde{\Gamma}_n + 2(P_n - P_0)P_0\left(\Gamma_n - \Gamma_0\right).\end{aligned}$$

By assumption, $\mathbb{U}_n\tilde{\Gamma}_n = o_{P_0}(n^{-1/2})$. The final term is $o_{P_0}(n^{-1/2})$ by the Donsker condition and the consistency condition [van der Vaart and Wellner, 1996]. By the second result of A.6 and the assumption that $K_n = o_{P_0}(n^{-1/2})$, this yields that

$$\psi_n - \psi_0 = 2(P_n - P_0)P_0\Gamma_0 + o_{P_0}(n^{-1/2}).$$

Multiplying both sides by $\sqrt{n}$, and applying the central limit theorem yields the result.

PROOF (PROOF OF 3). We have that

$$P_0^n\left\{n\psi_n \le \hat{q}_{1-\alpha}^{ub}\right\} = P_0^n\left\{\frac{\sqrt{n}(\psi_n - \psi_0)}{\sigma_0} \le \frac{\hat{q}_{1-\alpha}^{ub}n^{-1/2} - \sqrt{n}\psi_0}{\sigma_0}\right\}$$

Fix $0 < \epsilon < \psi_0$. The right-hand side is equal to

$$\begin{aligned}&P_0^n\left\{\frac{\sqrt{n}(\psi_n - \psi_0)}{\sigma_0} \le \frac{\hat{q}_{1-\alpha}^{ub}n^{-1/2} - \sqrt{n}\psi_0}{\sigma_0} \text{ and } \hat{q}_{1-\alpha}^{ub}n^{-1} \le \epsilon\right\} + o(1) \\ &\le P_0^n\left\{\frac{\sqrt{n}(\psi_n - \psi_0)}{\sigma_0} \le \frac{\sqrt{n}(\epsilon - \psi_0)}{\sigma_0} \text{ and } \hat{q}_{1-\alpha}^{ub}n^{-1} \le \epsilon\right\} + o(1) \\ &\le P_0^n\left\{\frac{\sqrt{n}(\psi_n - \psi_0)}{\sigma_0} \le \frac{\sqrt{n}(\epsilon - \psi_0)}{\sigma_0}\right\} + o(1) = Pr\left\{Z \le \frac{\sqrt{n}(\epsilon - \psi_0)}{\sigma_0}\right\} + o(1),\end{aligned}$$

where $Z \sim N(0,1)$. The final equality holds by Theorem 6 and the well known result about the uniform convergence of distribution functions at continuity points when random variables converge in distribution [see, e.g., Theorem 5.6 in Boos and Stefanski, 2013]. The result follows by noting that $(\epsilon - \psi_0)/\sigma_0$ is negative and that $\lim_{z\to-\infty} Pr(Z \le z) = 0$.

References

D D Boos and L A Stefanski. *Essential Statistical Inference: Theory and Methods*. Springer, Berlin Heidelberg New York, 2013. ISBN 978-1-4614-4817-4.

M Carone, I Díaz, and M J van der Laan. Higher-order Targeted Minimum Loss-based Estimation. 2014.

C M Fortuin, P W Kasteleyn, and J Ginibre. Correlation inequalities on some partially ordered sets. *Commun. Math. Phys.*, 22(2):89–103, 1971.

A Gretton, K M Borgwardt, M J Rasch, B Schölkopf, and A Smola. A kernel two-sample test. *J. Mach. Learn. Res.*, 13(1):723–773, 2012.

D Nolan and D Pollard. U-processes: rates of convergence. *Ann. Statist.*, 15(2):780–799, 1987.

D Nolan and D Pollard. Functional Limit Theorems for U-Processes. *Ann. Probab.*, 16(3): 1291–1298, 1988. ISSN 0091-1798. doi: 10.1214/aop/1176991691.

R J Serfling. *Approximation Theorems of Mathematical Statistics*, volume 37. 1980. ISBN 0471219274. doi: 10.2307/2530199.

A W van der Vaart. *Asymptotic statistics*. Cambridge University Press, New York, 1998.

A W van der Vaart and J A Wellner. *Weak convergence and empirical processes*. Springer, Berlin Heidelberg New York, 1996.