Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

SUPPLEMENT TO: PERMUTATION *P*-VALUE APPROXIMATION VIA GENERALIZED STOLARSKY INVARIANCE

By Hera Y. He, E-mail: hera.yu.he@gmail.com Kinjal Basu, E-mail: kbasu@linkedin.com Qingyuan Zhao, E-mail: qyzhao@wharton.upenn.edu Art B. Owen E-mail: owen@stanford.edu

Outline. This is an online supplement to the article "Permutation p-value approximation via generalized Stolarsky invariance". The section numbers pick up where the main article left off. Section 11 contains all but the shortest proofs of results in the main document. Section 12 investigates the effect of unbalanced sample sizes on the moments of the reference disributions. Section 13 analyzes the computational cost of computing $\hat{p}_2$ and its reference variance. Section 14 describes how to get the data we used.

11. Proofs.

Here we collect up the longer proofs. They appear in the same order that the corresponding lemmas and theorems appear in the main article.

11.1. *Proof of Lemma 1.*

Proof. Let $\boldsymbol{z} \sim \mathbf{U}(\mathbb{S}^d)$. Then $V_2(u;t,d) = \sigma_d(C_2(\boldsymbol{x},\boldsymbol{y};t)) = \Pr(\boldsymbol{z} \in C_2(\boldsymbol{x},\boldsymbol{y};t))$. If $u=1$ then $\boldsymbol{x}=\boldsymbol{y}$ and so $C_2(\boldsymbol{x},\boldsymbol{y};t) = C(\boldsymbol{x};t)$. For $u<1$, we decompose $\boldsymbol{y}$ and $\boldsymbol{z}$ with respect to $\boldsymbol{x}$, via $\boldsymbol{z} = s\boldsymbol{x} + \sqrt{1-s^2}\boldsymbol{z}^*$ and $\boldsymbol{y} = u\boldsymbol{x} + \sqrt{1-u^2}\boldsymbol{y}^*$. Now

$$
\begin{aligned}
V_2(u;t,d) &= \int_{\mathbb{S}^d} \mathbf{1}(\langle \boldsymbol{x},\boldsymbol{z}\rangle \geqslant t)\mathbf{1}(\langle \boldsymbol{y},\boldsymbol{z}\rangle \geqslant t)\,\mathrm{d}\sigma(\boldsymbol{z}) \\
&= \int_{-1}^{1} \mathbf{1}(s \geqslant t)\frac{\omega_{d-1}}{\omega_d}(1-s^2)^{\frac{d}{2}-1} \\
&\quad \times \int_{\mathbb{S}^{d-1}} \mathbf{1}(su + \sqrt{1-s^2}\sqrt{1-u^2}\,\langle \boldsymbol{y}^*,\boldsymbol{z}^*\rangle \geqslant t)\,\mathrm{d}\sigma_{d-1}(\boldsymbol{z}^*)\,\mathrm{d}s.
\end{aligned}
$$

If $u>-1$ then this reduces to (2.3). For $u=-1$ we get

$$
V_2(u;t,d) = \frac{\omega_{d-1}}{\omega_d}\int_{-1}^{1} \mathbf{1}(s \geqslant t)\mathbf{1}(-s \geqslant t)(1-s^2)^{\frac{d}{2}-1}\,\mathrm{d}s.
$$

which reduces to (2.4). □

imsart-aos ver. 2014/10/16 file: stolarsky-supplement.tex date: January 31, 2018

11.2. *Proof of Theorem 2 (Limiting invariance).* Here we show that taking limits as $\boldsymbol{\epsilon}$ goes to zero in the formula of Brauchart and Dick (2013) proves Theorem 2. We use three lemmas, one for each term in Theorem 1. We use $\widetilde{\lim}_{\boldsymbol{\epsilon}\to 0}$ as a shorthand for $\lim_{\epsilon_1\to 0^+}\lim_{\epsilon_2\to 0^+}$.

LEMMA 3. *Let $v_{\boldsymbol{\epsilon}}$ be defined as in (3.3). Then for $\hat\rho\in[-1,1)$,*

$$\widetilde{\lim}_{\boldsymbol{\epsilon}\to 0}\int_{-1}^{1} v_{\boldsymbol{\epsilon}}(t)\int_{\mathbb{S}^d}\Bigl|\sigma_d(C(\boldsymbol{z};t))-\frac{1}{N}\sum_{k=0}^{N-1}\mathbf{1}_{C(\boldsymbol{z};t)}(\boldsymbol{x}_k)\Bigr|^2\,\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}t$$
$$=\int_{\mathbb{S}^d}|p(\boldsymbol{z},\hat\rho)-\hat p(\hat\rho)|^2\,\mathrm{d}\sigma_d(\boldsymbol{z}).$$

PROOF. Substituting $v_{\boldsymbol{\epsilon}}$ we get

$$\int_{-1}^{1}\Bigl(\epsilon_2+\frac{1}{\epsilon_1}\mathbf{1}(\hat\rho\leqslant t\leqslant\hat\rho+\epsilon_1)\Bigr)\int_{\mathbb{S}^d}(\hat p(t)-p(\boldsymbol{z},t))^2\,\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}t$$
$$\to\quad\frac{1}{\epsilon_1}\int_{\hat\rho}^{\hat\rho+\epsilon_1}\int_{\mathbb{S}^d}(\hat p(t)-p(\boldsymbol{z},t))^2\,\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}t,\quad\text{as }\epsilon_2\to 0^+$$
$$\to\quad\int_{\mathbb{S}^d}(\hat p(\hat\rho)-p(\boldsymbol{z},\hat\rho))^2\,\mathrm{d}\sigma_d(\boldsymbol{z}),\quad\text{as }\epsilon_1\to 0^+.\qquad\square$$

LEMMA 4. *Let $v_{\boldsymbol{\epsilon}}$ be as in (3.3) with $\hat\rho\in[-1,1)$, and let $K_{v_{\boldsymbol{\epsilon}}}$ be given by (3.2). Then for any $\boldsymbol{x},\boldsymbol{x}'\in\mathbb{S}^d$,*

$$\widetilde{\lim}_{\boldsymbol{\epsilon}\to 0}K_{v_{\boldsymbol{\epsilon}}}(\boldsymbol{x},\boldsymbol{x}')=\sigma_d(C(\boldsymbol{x};\hat\rho)\cap C(\boldsymbol{x}';\hat\rho)).$$

PROOF. The argument is essentially the same as for Lemma 3. $\square$

LEMMA 5. *Let $v_{\boldsymbol{\epsilon}}$ be as in (3.3) with $\hat\rho\in[-1,1)$, and let $K_{v_{\boldsymbol{\epsilon}}}$ be given by (3.2). Then*

$$\widetilde{\lim}_{\boldsymbol{\epsilon}\to 0}\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}K_{v_{\boldsymbol{\epsilon}}}(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{x})\,\mathrm{d}\sigma_d(\boldsymbol{y})=\hat p_1(\hat\rho)^2.\tag{11.1}$$

PROOF. For any $\boldsymbol{x},\boldsymbol{y}\in\mathbb{S}^d$, the kernel $K_{v_{\boldsymbol{\epsilon}}}(\boldsymbol{x},\boldsymbol{y})$ is nonnegative and upper bounded by a constant. Therefore we can take our limit operations inside the double integral over $\boldsymbol{x}$ and $\boldsymbol{y}$. Now $\widetilde{\lim}_{\boldsymbol{\epsilon}\to 0}K_{v_{\boldsymbol{\epsilon}}}(\boldsymbol{x},\boldsymbol{y})=\int_{\mathbb{S}^d}\mathbf{1}_{C(\boldsymbol{z};\hat\rho)}(\boldsymbol{x})\mathbf{1}_{C(\boldsymbol{z};\hat\rho)}(\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{z})$. Therefore the limit in (11.1) is

$$\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\mathbf{1}_{C(\boldsymbol{z};\hat\rho)}(\boldsymbol{x})\mathbf{1}_{C(\boldsymbol{z};\hat\rho)}(\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}\sigma_d(\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{x})=\hat p_1(\hat\rho)^2$$

after changing the order of the integrals. $\square$

Theorem 2 *Let $\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_{N-1} \in \mathbb{S}^d$ and $\hat{\rho} \in [-1,1]$. Then*

$$\int_{\mathbb{S}^d} |p(\boldsymbol{z}, \hat{\rho}) - \hat{p}_1(\hat{\rho})|^2 \,\mathrm{d}\sigma_d(\boldsymbol{z}) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{\ell=0}^{N-1} \sigma_d(C(\boldsymbol{x}_k; \hat{\rho}) \cap C(\boldsymbol{x}_\ell; \hat{\rho})) - \hat{p}_1(\hat{\rho})^2.$$

Proof. Theorem 1 gives us an identity and applying Lemmas 3, 4 and 5 to both sides of it establishes (3.4) for $\hat{\rho} \in [-1,1)$. For $\hat{\rho} = 1$ we get the answer by replacing $v_{\boldsymbol{\epsilon}}$ by $\epsilon_2 + (1/\epsilon_1)\mathbf{1}_{1-\epsilon_1 \leqslant t \leqslant 1}$ in the lemmas. Replacing $\boldsymbol{z}$ by $\boldsymbol{y}$ and $\hat{\rho}$ by t above gives the version in the main body of the article. □

11.3. *Proof of Lemma 3 (Double inclusion for reference distribution 2).*

Proof. We split the proof into four cases and prove them individually. Recall that $P_2(u_1, u_2, u_3, \tilde{\rho}, \hat{\rho}) = \int_{\mathbb{S}^{d-1}} \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_1 \rangle \geqslant \hat{\rho}) \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_2 \rangle \geqslant \hat{\rho}) \,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*)$ where $\boldsymbol{y} = \tilde{\rho}\boldsymbol{x}_c + \sqrt{1-\tilde{\rho}^2}\boldsymbol{y}^*$.

Case 1. $\boldsymbol{x}_1 = \boldsymbol{x}_2 = \boldsymbol{x}_c$, i.e., $r_1 = r_2 = r_3 = 0$.

$$P_2(1,1,1,\tilde{\rho},\hat{\rho}) = \int_{\mathbb{S}^{d-1}} \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_c \rangle \geqslant \hat{\rho}) \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_c \rangle \geqslant \hat{\rho}) \,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*) = \mathbf{1}(\tilde{\rho} \geqslant \hat{\rho}).$$

Case 2. $\boldsymbol{x}_1 = \boldsymbol{x}_c \neq \boldsymbol{x}_2$, i.e., $r_1 = 0, r_2 > 0, r_3 > 0$.

$$\begin{aligned} P_2(1,u_2,u_2,\tilde{\rho},\hat{\rho}) &= \int_{\mathbb{S}^{d-1}} \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_c \rangle \geqslant \hat{\rho}) \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_2 \rangle \geqslant \hat{\rho}) \,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*) \\ &= \mathbf{1}(\tilde{\rho} \geqslant \hat{\rho}) \int_{\mathbb{S}^{d-1}} \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_2 \rangle \geqslant \hat{\rho}) \,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*) \\ &= \mathbf{1}(\tilde{\rho} \geqslant \hat{\rho}) P_1(u_2, \tilde{\rho}, \hat{\rho}) \end{aligned}$$

where the last step uses Lemma 2.

Case 3. $\boldsymbol{x}_1 = \boldsymbol{x}_2 \neq \boldsymbol{x}_c$, i.e., $r_1 = r_2 > 0 = r_3$.

$$\begin{aligned} P_2(u_2,u_2,1,\tilde{\rho},\hat{\rho}) &= \int_{\mathbb{S}^{d-1}} \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_1 \rangle \geqslant \hat{\rho}) \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_2 \rangle \geqslant \hat{\rho}) \,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*) \\ &= \int_{\mathbb{S}^{d-1}} \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_2 \rangle \geqslant \hat{\rho}) \,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*) \\ &= P_1(u_2, \tilde{\rho}, \hat{\rho}). \end{aligned}$$

Case 4. $\boldsymbol{x}_1 \neq \boldsymbol{x}_2 \neq \boldsymbol{x}_c \neq \boldsymbol{x}_1$, i.e., $r_1, r_2, r_3 > 0$. We split this case into subcases. First we assume $u_2 = -1$, so

$$\begin{aligned} P_2(u_1,u_2,u_3,\tilde{\rho},\hat{\rho}) &= \int_{\mathbb{S}^{d-1}} \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_1 \rangle \geqslant \hat{\rho}) \mathbf{1}(\langle \boldsymbol{y}, -\boldsymbol{x}_c \rangle \geqslant \hat{\rho}) \,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*) \\ &= \mathbf{1}(-\tilde{\rho} \geqslant \hat{\rho}) \int_{\mathbb{S}^{d-1}} \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_1 \rangle \geqslant \hat{\rho}) \,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*) \\ &= \mathbf{1}(-\tilde{\rho} \geqslant \hat{\rho}) P_1(u_1, \tilde{\rho}, \hat{\rho}). \end{aligned}$$

Similarly if $u_1 = -1$, then

$$P_2(u_1, u_2, u_3, \tilde{\rho}, \hat{\rho}) = \mathbf{1}(-\tilde{\rho} \geqslant \hat{\rho}) P_1(u_2, \tilde{\rho}, \hat{\rho}).$$

Finally we assume $u_1 > -1$ and $u_2 > -1$, so now $|u_1| < 1$ and $|u_2| < 1$. Recall the decompositions $\boldsymbol{x}_j = u_j \boldsymbol{c}_c + \sqrt{1 - u_j^2}\boldsymbol{x}_j^*$ for $j = 1, 2$ and introduce further decompositions of $\boldsymbol{y}^*$ and $\boldsymbol{x}_2^*$ with respect to $\boldsymbol{x}_1^*$: $\boldsymbol{y}^* = t\boldsymbol{x}_1^* + \sqrt{1-t^2}\boldsymbol{y}^{**}$ and $\boldsymbol{x}_2^* = u_3^*\boldsymbol{x}_1^* + \sqrt{1-u_3^*}\boldsymbol{x}_2^{**}$. The residuals $\boldsymbol{y}^{**}$ and $\boldsymbol{x}_2^{**}$ belong to a subset of $\mathbb{S}^d$ that is isomorphic to $\mathbb{S}^{d-2}$. Now we have

$$\begin{aligned}
&P_2(u_1, u_2, u_3, \tilde{\rho}, \hat{\rho}) \\
&= \int_{\mathbb{S}^{d-1}} \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_1\rangle \geqslant \hat{\rho}) \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_2\rangle \geqslant \hat{\rho}) \,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*) \\
&= \int_{-1}^{1} \frac{\omega_{d-2}}{\omega_{d-1}} (1-t^2)^{\frac{d-1}{2}-1} \int_{\mathbb{S}^{d-2}} \mathbf{1}\Big(\tilde{\rho} u_1 + \sqrt{1-\tilde{\rho}^2}\sqrt{1-u_1^2}t \geqslant \hat{\rho}\Big) \\
&\quad \times \mathbf{1}\Big(\tilde{\rho} u_2 + \sqrt{1-\tilde{\rho}^2}\sqrt{1-u_2^2}(tu_3^* + \sqrt{1-t^2}\sqrt{1-u_3^{*2}}\,\langle \boldsymbol{y}^{**}, \boldsymbol{x}_2^{**}\rangle) \geqslant \hat{\rho}\Big) \\
&\quad \times \,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^{**}) \,\mathrm{d}t \\
&= \int_{-1}^{1} \frac{\omega_{d-2}}{\omega_{d-1}} (1-t^2)^{\frac{d-1}{2}-1} \mathbf{1}\Big(t \geqslant \frac{\hat{\rho} - \tilde{\rho} u_1}{\sqrt{1-\tilde{\rho}^2}\sqrt{1-u_1^2}}\Big) \\
&\quad \times \int_{\mathbb{S}^{d-2}} \mathbf{1}\Big(\tilde{\rho} u_2 + \sqrt{1-\tilde{\rho}^2}\sqrt{1-u_2^2}(tu_3^* + \sqrt{1-t^2}\sqrt{1-u_3^{*2}}\,\langle \boldsymbol{y}^{**}, \boldsymbol{x}_2^{**}\rangle) \geqslant \hat{\rho}\Big) \\
&\quad \times \,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^{**}) \,\mathrm{d}t \\
&= \begin{cases}
\mathbf{1}(\tilde{\rho} u_1 \geqslant \hat{\rho}) \mathbf{1}(\tilde{\rho} u_2 \geqslant \hat{\rho}), & \tilde{\rho} = \pm 1 \\
\int_{-1}^{1} \frac{\omega_{d-2}}{\omega_{d-1}} (1-t^2)^{\frac{d-1}{2}-1} \mathbf{1}(t \geqslant \rho_1) \mathbf{1}(tu_3^* \geqslant \rho_2) \,\mathrm{d}t, & \tilde{\rho} \neq \pm 1, u_3^* = \pm 1 \\
\int_{-1}^{1} \frac{\omega_{d-2}}{\omega_{d-1}} (1-t^2)^{\frac{d-1}{2}-1} \mathbf{1}(t \geqslant \rho_1) \sigma_{d-2}(C(\boldsymbol{x}_2^{**}, \frac{\rho_2 - tu_3^*}{\sqrt{1-t^2}\sqrt{1-u_3^{*2}}})) \,\mathrm{d}t, & \tilde{\rho} \neq \pm 1, |u_3^*| < 1
\end{cases}
\end{aligned}$$

where u_3^*, ρ_1, ρ_2 are defined in (4.7). Hence, the result follows. □

11.4. *Proof of Theorem 5 (Second moment under reference distribution 2).*

PROOF. Without loss of generality we relabel the values $\boldsymbol{x}_k$ so that $c = 0$. If the original $\boldsymbol{x}_c$ was not $\boldsymbol{x}_0$, that choice is captured by the number $\widetilde{\rho} \neq \hat{\rho}$. The second moment is

$$\mathbb{E}(p(\boldsymbol{y}, \hat{\rho})^2) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{\ell=0}^{N-1} P_2(u_k, u_\ell, u_{k,\ell}, \tilde{\rho}, \hat{\rho}) \tag{11.2}$$

where $u_{k,\ell}$ is obtained via (2.1) from the swap distance $r_{k,\ell}$ between points $\boldsymbol{x}_k$ and $\boldsymbol{x}_\ell$. We will partition the sum in (11.2) into the same four cases as in the proof of Lemma 3.

Case 1, $\boldsymbol{x}_k = \boldsymbol{x}_\ell = \boldsymbol{x}_c$, i.e., $r_k = r_\ell = r_{k,\ell} = 0$. There is only one pair of $(\boldsymbol{x}_k, \boldsymbol{x}_\ell)$ for this condition. Hence, we get only one term corresponding to $P_2(1,1,1,\tilde\rho,\hat\rho) = \mathbf{1}(\tilde\rho \geqslant \hat\rho)$.

Case 2, $\boldsymbol{x}_k = \boldsymbol{x}_c \neq \boldsymbol{x}_\ell$, i.e., $r_k = 0, r_\ell = r_{k,\ell} > 0$. Consider all pairs of $(\boldsymbol{x}_k, \boldsymbol{x}_\ell)$ that satisfy this condition and let K_2 denote their total contribution to (11.2). Then

$$\begin{aligned} K_2 &= 2\sum_{\ell=1}^{N-1} \int_{\mathbb{S}^{d-1}} \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_c\rangle \geqslant \hat\rho)\mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_\ell\rangle \geqslant \hat\rho)\,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*) \\ &= 2\sum_{r=1}^{\mathrm{m}} \binom{m_0}{r}\binom{m_1}{r} P_2(1, u(r), u(r), \tilde\rho, \hat\rho). \end{aligned}$$

Case 3, $\boldsymbol{x}_k = \boldsymbol{x}_\ell \neq \boldsymbol{x}_c$, i.e., $r_k = r_\ell > 0 = r_{k,\ell}$. The contribution from terms of this form is

$$K_3 = \sum_{k=1}^{N-1} \int_{\mathbb{S}^{d-1}} \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_k\rangle \geqslant \hat\rho)\,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*) = \sum_{r=1}^{\mathrm{m}} \binom{m_0}{r}\binom{m_1}{r} P_1(u(r), \tilde\rho, \hat\rho).$$

Case 4, $\boldsymbol{x}_k \neq \boldsymbol{x}_\ell \neq \boldsymbol{x}_c$, i.e., $r_k, r_\ell, r_{k,\ell} > 0$. The contribution of these cases to the sum is

$$\begin{aligned} K_4 &= \sum_{k=1}^{N-1}\sum_{\ell=1}^{N-1} \mathbf{1}(\ell \neq k) \int_{\mathbb{S}^{d-1}} \mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_i\rangle \geqslant \hat\rho)\mathbf{1}(\langle \boldsymbol{y}, \boldsymbol{x}_j\rangle \geqslant \hat\rho)\,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*) \\ &= \sum_{r_k\in R}\sum_{r_\ell\in R}\sum_{r_{k,\ell}\in R_3(\boldsymbol{r})} c(r_k, r_\ell, r_{k,\ell}) P_2(u_1, u_2, u_3, \tilde\rho, \hat\rho). \end{aligned}$$

Then the second moment is $(\mathbf{1}(\tilde\rho \geqslant \hat\rho) + K_2 + K_3 + K_4)/N^2$. □

11.5. *Proof of Theorem 6 (Location weighted invariance)* .

Proof. We follow the technique in Brauchart and Dick (2013). Like them, we use basic properties of reproducing kernel Hilbert spaces. References to Aronszajn (1950) below are to pages 343–344 of their Section 2. We begin by showing that $K_{v,h,\boldsymbol{x}'}$ as defined in (5.2) is a reproducing kernel. First, $K_{v,h,\boldsymbol{x}'}$ is symmetric: $K_{v,h,\boldsymbol{x}'}(\boldsymbol{x},\boldsymbol{y}) = K_{v,h,\boldsymbol{x}'}(\boldsymbol{y},\boldsymbol{x})$. Next, choose

$a_0, \dots, a_{N-1} \in \mathbb{R}$ and $\boldsymbol{x}_0, \dots, \boldsymbol{x}_{N-1} \in \mathbb{S}^d$. Then $\sum_{k,\ell=0}^{N-1} a_k a_\ell K_{v,h,\boldsymbol{x}'}(\boldsymbol{x}_k, \boldsymbol{x}_\ell)$ equals

$$\begin{aligned}&\int_{-1}^{1}\int_{\mathbb{S}^d}\sum_{k,\ell=0}^{N-1} a_k a_\ell v(t) h(\langle \boldsymbol{z}, \boldsymbol{x}'\rangle)\mathbf{1}_{C(\boldsymbol{z};t)}(\boldsymbol{x}_k)\mathbf{1}_{C(\boldsymbol{z};t)}(\boldsymbol{x}_\ell)\,\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}t\\ &=\int_{-1}^{1}\int_{\mathbb{S}^d} v(t)h(\langle \boldsymbol{z}, \boldsymbol{x}'\rangle)\Bigl|\sum_{k=0}^{N-1} a_k \mathbf{1}_{C(\boldsymbol{z},t)}(\boldsymbol{x}_k)\Bigr|^2\,\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}t\end{aligned}$$

which is nonnegative. Thus $K_{v,h,\boldsymbol{x}'}$ is symmetric and positive definite, and so by Aronszajn (1950), $K_{v,h,\boldsymbol{x}'}$ is a reproducing kernel.

Aronszajn (1950) also shows that a reproducing kernel uniquely defines a Hilbert space of functions with a specific inner product. Let $\mathcal{H}_{v,h,\boldsymbol{x}'} = \mathcal{H}(K_{v,h,\boldsymbol{x}'}, \mathbb{S}^d)$ denote the corresponding reproducing kernel Hilbert space of functions $f : \mathbb{S}^d \to \mathbb{R}$ with reproducing kernel $K_{v,h,\boldsymbol{x}'}$.

We now consider functions $f_1, f_2 : \mathbb{S}^d \to \mathbb{R}$ which admit the representation

$$f_i(\boldsymbol{x}) = \int_{-1}^{1}\int_{\mathbb{S}^d} g_i(\boldsymbol{z};t)\mathbf{1}_{C(\boldsymbol{z};t)}(\boldsymbol{x})\,\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}t, \quad i = 1, 2 \tag{11.3}$$

for functions $g_i \in L_2(\mathbb{S}^d \times [-1, 1])$. For any fixed $\boldsymbol{y} \in \mathbb{S}^d$ the function $K_{v,h,\boldsymbol{x}'}(\cdot, \boldsymbol{y})$ has representation (11.3) via $g(\boldsymbol{z};t) = v(t)h(\langle \boldsymbol{z}, \boldsymbol{x}'\rangle)\mathbf{1}_{C(\boldsymbol{z},t)}(\boldsymbol{y})$.

For functions with representation (11.3), we define an inner product by

$$\langle f_1, f_2\rangle_{K_{v,h,\boldsymbol{x}'}} = \int_{-1}^{1}\frac{1}{v(t)}\int_{\mathbb{S}^d}\frac{1}{h(\langle \boldsymbol{z}, \boldsymbol{x}'\rangle)} g_1(\boldsymbol{z},t)g_2(\boldsymbol{z},t)\,\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}t. \tag{11.4}$$

For $\boldsymbol{y} \in \mathbb{S}^d$ and $f_1 \in \mathcal{H}_{v,h,\boldsymbol{x}'}$,

$$\begin{aligned}\bigl\langle f_1, K_{v,h,\boldsymbol{x}'}(\cdot,\boldsymbol{y})\bigr\rangle_{K_{v,h,\boldsymbol{x}'}} &= \int_{-1}^{1}\frac{1}{v(t)}\int_{\mathbb{S}^d}\frac{g_1(\boldsymbol{z};t)v(t)}{h(\langle \boldsymbol{z}, \boldsymbol{x}'\rangle)}h(\langle \boldsymbol{z}, \boldsymbol{x}'\rangle)\mathbf{1}_{C(\boldsymbol{z};t)}(\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}t\\ &= \int_{-1}^{1}\int_{\mathbb{S}^d} g_1(\boldsymbol{z},t)\mathbf{1}_{C(\boldsymbol{z},t)}(\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}t\\ &= f_1(\boldsymbol{y}),\end{aligned}$$

showing that the inner product (11.4) has the reproducing property. By Aronszajn (1950), the inner product in $\mathcal{H}_{v,h,\boldsymbol{x}'}$ is unique. Functions f_i satisfying (11.3) with $\langle f_i, f_i\rangle_{K_{v,h,\boldsymbol{x}'}} < \infty$ are in $\mathcal{H}_{v,h,\boldsymbol{x}'}$, and (11.4) is the unique inner product of $\mathcal{H}_{v,h,\boldsymbol{x}'}$.

We prove the theorem by equating two different forms of $\|\mathcal{R}(\mathcal{H}_{v,h,\boldsymbol{x}'};\cdot\,)\|_{K_{v,h,\boldsymbol{x}'}}$ where

$$\mathcal{R}(\mathcal{H}_{v,h,\boldsymbol{x}'};\cdot\,) = \int_{\mathbb{S}^d} K_{v,h,\boldsymbol{x}'}(\cdot\,,\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{y}) - \frac{1}{N}\sum_{k=0}^{N-1} K_{v,h,\boldsymbol{x}'}(\cdot\,,\boldsymbol{x}_k).$$

Although $\mathcal{R}(\mathcal{H}_{v,h,\boldsymbol{x}'};\cdot\,)$ depends on our specific points $\boldsymbol{x}_i$ we omit that from the notation. The reproducing property of $K_{v,h,\boldsymbol{x}'}$ yields

$$\big\langle K_{v,h,\boldsymbol{x}'}(\cdot\,,\boldsymbol{x}_k), K_{v,h,\boldsymbol{x}'}(\cdot\,,\boldsymbol{x}_\ell)\big\rangle_{K_{v,h,\boldsymbol{x}'}} = K_{v,h,\boldsymbol{x}'}(\boldsymbol{x}_k,\boldsymbol{x}_\ell)$$

from which it follows that

$$\begin{aligned}&\left\langle \int_{\mathbb{S}^d} K_{v,h,\boldsymbol{x}'}(\cdot\,,\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{y}), \int_{\mathbb{S}^d} K_{v,h,\boldsymbol{x}'}(\cdot\,,\boldsymbol{y}')\,\mathrm{d}\sigma_d(\boldsymbol{y}')\right\rangle_{K_{v,h,\boldsymbol{x}'}}\\ &= \int_{\mathbb{S}^d}\int_{\mathbb{S}^d} K_{v,h,\boldsymbol{x}'}(\boldsymbol{y},\boldsymbol{y}')\,\mathrm{d}\sigma_d(\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{y}').\end{aligned} \tag{11.5}$$

Using (11.5) and the linearity of the inner product, we have

$$\begin{aligned}&\big\langle \mathcal{R}(\mathcal{H}_{v,h,\boldsymbol{x}'};\cdot\,), \mathcal{R}(\mathcal{H}_{v,h,\boldsymbol{x}'};\cdot\,)\big\rangle_{K_{v,h,\boldsymbol{x}'}}\\ &= \int_{\mathbb{S}^d}\int_{\mathbb{S}^d} K_{v,h,\boldsymbol{x}'}(\boldsymbol{y},\boldsymbol{y}')\,\mathrm{d}\sigma_d(\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{y}') - \frac{2}{N}\sum_{k=0}^{N-1}\int_{\mathbb{S}^d} K_{v,h,\boldsymbol{x}'}(\boldsymbol{y},\boldsymbol{x}_k)\,\mathrm{d}\sigma_d(\boldsymbol{y})\\ &+ \frac{1}{N^2}\sum_{k,\ell=0}^{N-1} K_{v,h,\boldsymbol{x}'}(\boldsymbol{x}_k,\boldsymbol{x}_\ell).\end{aligned} \tag{11.6}$$

For our second form of $\|\mathcal{R}(\mathcal{H}_{v,h,\boldsymbol{x}'};\cdot\,)\|_{K_{v,h,\boldsymbol{x}'}}$, we write

$$\begin{aligned}&\mathcal{R}(\mathcal{H}_{v,h,\boldsymbol{x}'};\cdot\,)\\ &= \int_{\mathbb{S}^d} K_{v,h,\boldsymbol{x}'}(\cdot\,,\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{y}) - \frac{1}{N}\sum_{k=0}^{N-1} K_{v,h,\boldsymbol{x}'}(\cdot\,,\boldsymbol{x}_k)\\ &= \int_{-1}^{1} v(t)\int_{\mathbb{S}^d} \mathbf{1}_{C(\boldsymbol{z};t)}(\boldsymbol{x})h(\langle \boldsymbol{z},\boldsymbol{x}\rangle)\left[\int_{\mathbb{S}^d} \mathbf{1}_{C(\boldsymbol{z},t)}(\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{y})\,\mathrm{d}t - \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{1}_{C(\boldsymbol{z},t)}(\boldsymbol{x}_k)\right]\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}t\\ &= \int_{-1}^{1} v(t)\int_{\mathbb{S}^d} \mathbf{1}_{C(\boldsymbol{z},t)}(x)h(\langle \boldsymbol{z},\boldsymbol{x}\rangle)\left[\sigma_d(C(\boldsymbol{z},t)) - \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{1}_{C(\boldsymbol{z},t)}(\boldsymbol{x}_k)\right]\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}t.\end{aligned}$$

Hence using the definition of the inner product $\langle\cdot,\cdot\rangle_{K_{v,h,\boldsymbol{x}'}}$, we have

(11.7)
$$\begin{aligned}&\big\langle \mathcal{R}(\mathcal{H}_{v,h,\boldsymbol{x}'};\boldsymbol{x}),\mathcal{R}(\mathcal{H}_{v,h,\boldsymbol{x}'};\boldsymbol{x})\big\rangle_{K_{v,h,\boldsymbol{x}'}}\\ &=\int_{-1}^{1} v(t)\int_{\mathbb{S}^d} h(\langle \boldsymbol{z},\boldsymbol{x}\rangle)\bigg|\sigma_d(C(\boldsymbol{x},t))-\frac{1}{N}\sum_{k=0}^{N-1}\mathbf{1}_{C(\boldsymbol{x};t)}(\boldsymbol{x}_k)\bigg|^2\,\mathrm{d}\sigma_d(\boldsymbol{x})\,\mathrm{d}t.\end{aligned}$$

Combining equations (11.6) and (11.7), we have the generalized location-weighted version of the Stolarsky invariance principle. □

11.6. *Proof of Theorem 7 (Spatially weighed invariance)* . As in Section 11.2, $\widetilde{\lim}_{\boldsymbol{\epsilon}\to 0}$ means $\lim_{\epsilon_1\to 0^+}\lim_{\epsilon_2\to 0^+}$ and similarly $\widetilde{\lim}_{\boldsymbol{\eta}\to 0}$ denotes $\lim_{\eta_1\to 0^+}\lim_{\eta_2\to 0^+}$. We prove a series of lemmas first.

LEMMA 6. *For $v_{\boldsymbol{\epsilon}}(\cdot)$ and $h_{\boldsymbol{\eta}}(\cdot)$ defined by equations (3.3) and (5.3),*

$$\begin{aligned}&\widetilde{\lim}_{\boldsymbol{\eta}\to 0}\widetilde{\lim}_{\boldsymbol{\epsilon}\to 0}\int_{-1}^{1} v_{\boldsymbol{\epsilon}}(t)\int_{\mathbb{S}^d} h_{\boldsymbol{\eta}}(\langle \boldsymbol{z},\boldsymbol{x}_c\rangle)\bigg|\sigma_d(C(\boldsymbol{z},t))-\frac{1}{N}\sum_{k=0}^{N-1}\mathbf{1}_{C(\boldsymbol{z};t)}(\boldsymbol{x}_k)\bigg|^2\,\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}t\\ &=\int_{\mathbb{S}^{d-1}}|p(\tilde\rho\boldsymbol{x}_c+\sqrt{1-\tilde\rho^2}\boldsymbol{y}^*,\hat\rho)-\hat p_1(\hat\rho)|^2\,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*),\end{aligned}$$

where $\hat p_1(\hat\rho)=\sigma_d(C(\boldsymbol{y};\hat\rho))$.

PROOF. This proof is similar to the others. First we take the limit $\boldsymbol{\epsilon}\to 0$ yielding

$$\widetilde{\lim}_{\boldsymbol{\eta}\to 0}\int_{\mathbb{S}^d} h_{\boldsymbol{\eta}}(\langle \boldsymbol{z},\boldsymbol{x}_c\rangle)\bigg|\sigma_d(C(\boldsymbol{z},\hat\rho))-\frac{1}{N}\sum_{k=1}^{N}\mathbf{1}_{C(\boldsymbol{z};\hat\rho)}(\boldsymbol{x}_k)\bigg|^2\,\mathrm{d}\sigma_d(\boldsymbol{z}).$$

Making the decomposition $\boldsymbol{z}=s\boldsymbol{x}_c+\sqrt{1-s^2}\boldsymbol{z}^*$ gives

$$\begin{aligned}&\widetilde{\lim}_{\boldsymbol{\eta}\to 0}\int_{-1}^{1}\int_{\mathbb{S}^{d-1}}\frac{\omega_{d-1}}{\omega_d}(1-s^2)^{d/2-1}h_{\boldsymbol{\eta}}(s)\times\\ &\qquad\bigg|\sigma_d(C(s\boldsymbol{x}_c+\sqrt{1-s^2}\boldsymbol{z}^*,\hat\rho))-\frac{1}{N}\sum_{k=1}^{N}\mathbf{1}_{C(s\boldsymbol{x}_c+\sqrt{1-s^2}\boldsymbol{z}^*;\hat\rho)}(\boldsymbol{x}_k)\bigg|^2\,\mathrm{d}\sigma_{d-1}(\boldsymbol{z}^*)\,\mathrm{d}s\\ &=\int_{\mathbb{S}^{d-1}}|p(\hat\rho\boldsymbol{x}_c+\sqrt{1-\hat\rho^2}\boldsymbol{y}^*,\hat\rho)-\hat p_1(\hat\rho)|^2\,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*).\end{aligned}$$
□

LEMMA 7. *For $v_{\boldsymbol{\epsilon}}(\cdot)$ and $h_{\boldsymbol{\eta}}(\cdot)$ defined by equations (3.3) and (5.3),*

$$\widetilde{\lim}_{\boldsymbol{\eta}\to 0}\widetilde{\lim}_{\boldsymbol{\epsilon}\to 0}\frac{1}{N^2}\sum_{k,\ell=0}^{N-1}K_{v_{\boldsymbol{\epsilon}},h_{\boldsymbol{\eta}},\boldsymbol{x}_c}(\boldsymbol{x}_k,\boldsymbol{x}_\ell)$$
$$=\frac{1}{N^2}\sum_{k,\ell=0}^{N-1}\int_{\mathbb{S}^{d-1}}\mathbf{1}(\langle\boldsymbol{y},\boldsymbol{x}_k\rangle\geqslant\hat{\rho})\mathbf{1}(\langle\boldsymbol{y},\boldsymbol{x}_\ell\rangle\geqslant\hat{\rho})\,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*).$$

PROOF. First, $\widetilde{\lim}_{\boldsymbol{\eta}\to 0}\widetilde{\lim}_{\boldsymbol{\epsilon}\to 0}N^{-2}\sum_{k,\ell=0}^{N-1}K_{v_{\boldsymbol{\epsilon}},h_{\boldsymbol{\eta}},\boldsymbol{x}_c}(\boldsymbol{x}_k,\boldsymbol{x}_\ell)$ equals

$$\frac{1}{N^2}\sum_{k,\ell=0}^{N-1}\widetilde{\lim}_{\boldsymbol{\eta}\to 0}\int_{\mathbb{S}^d}h_{\boldsymbol{\eta}}(\langle\boldsymbol{z},\boldsymbol{x}_c\rangle)\mathbf{1}_{C(\boldsymbol{z};\hat{\rho})}(\boldsymbol{x}_k)\mathbf{1}_{C(\boldsymbol{z};\hat{\rho})}(\boldsymbol{x}_\ell)\,\mathrm{d}\sigma_d(\boldsymbol{z}).$$

Taking the decomposition of $\boldsymbol{z}$ with respect to $\boldsymbol{x}_c$ yields $\boldsymbol{z}=s\boldsymbol{x}_c+\sqrt{1-s^2}\boldsymbol{y}^*$ and then we have

$$\frac{1}{N^2}\sum_{k,\ell=0}^{N-1}\widetilde{\lim}_{\boldsymbol{\eta}\to 0}\int_{-1}^{1}\frac{\omega_{d-1}}{\omega_d}(1-s^2)^{d/2-1}h_{\boldsymbol{\eta}}(s)\int_{\mathbb{S}^{d-1}}\mathbf{1}_{C(\boldsymbol{z};\hat{\rho})}(\boldsymbol{x}_k)\mathbf{1}_{C(\boldsymbol{z};\hat{\rho})}(\boldsymbol{x}_\ell)\,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*)$$
$$=\frac{1}{N^2}\sum_{k,\ell=0}^{N-1}\int_{\mathbb{S}^{d-1}}\mathbf{1}(\langle\boldsymbol{y},\boldsymbol{x}_k\rangle\geqslant\hat{\rho})\mathbf{1}(\langle\boldsymbol{y},\boldsymbol{x}_\ell\rangle\geqslant\hat{\rho})\,\mathrm{d}\sigma_{d-1}(\boldsymbol{y}^*).\qquad\square$$

LEMMA 8. *For $v_{\boldsymbol{\epsilon}}(\cdot)$ and $h_{\boldsymbol{\eta}}(\cdot)$ defined by equations (3.3) and (5.3),*

$$\widetilde{\lim}_{\boldsymbol{\eta}\to 0}\widetilde{\lim}_{\boldsymbol{\epsilon}\to 0}\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}K_{v_{\boldsymbol{\epsilon}},h_{\boldsymbol{\eta}},\boldsymbol{x}_c}(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{x})\,\mathrm{d}\sigma_d(\boldsymbol{y})=\hat{p}_1(\hat{\rho})^2.$$

PROOF. Because $K_{v_{\boldsymbol{\epsilon}},h_{\boldsymbol{\eta}},\boldsymbol{x}_c}$ is nonnegative and uniformly bounded we may take the limit over $\boldsymbol{\epsilon}$ inside the integrals. Now

$$\widetilde{\lim}_{\boldsymbol{\epsilon}\to 0}K_{v_{\boldsymbol{\epsilon}},h_{\boldsymbol{\eta}},\boldsymbol{x}_c}(\boldsymbol{x},\boldsymbol{y})=\int_{\mathbb{S}^d}h_{\boldsymbol{\eta}}(\langle\boldsymbol{z},\boldsymbol{x}_c\rangle)\mathbf{1}_{C(\boldsymbol{z};\hat{\rho})}(\boldsymbol{x})\mathbf{1}_{C(\boldsymbol{z};\hat{\rho})}(\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{z}),$$

and the limit becomes

$$\widetilde{\lim}_{\boldsymbol{\eta}\to 0}\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}h_{\boldsymbol{\eta}}(\langle\boldsymbol{z},\boldsymbol{x}_c\rangle)\mathbf{1}_{C(\boldsymbol{z};\hat{\rho})}(\boldsymbol{x})\mathbf{1}_{C(\boldsymbol{z};\hat{\rho})}(\boldsymbol{y})\,\mathrm{d}\sigma_d(\boldsymbol{z})\,\mathrm{d}\sigma_d(\boldsymbol{x})\,\mathrm{d}\sigma_d(\boldsymbol{y}).$$

Integrating over $\boldsymbol{z}$ last we get $\widetilde{\lim}_{\boldsymbol{\eta}\to 0}\int_{\mathbb{S}^d}h_{\boldsymbol{\eta}}(\langle\boldsymbol{z},\boldsymbol{x}_c\rangle)\hat{p}_1^2(\hat{\rho})\,\mathrm{d}\boldsymbol{z}=\hat{p}_1^2(\hat{\rho})$. $\square$

LEMMA 9. *Under reference distribution 2*

$$\widetilde{\lim}_{\boldsymbol{\eta}\to 0}\widetilde{\lim}_{\boldsymbol{\epsilon}\to 0}\frac{2}{N}\sum_{k=0}^{N-1}\int_{\mathbb{S}^d} K_{v_{\boldsymbol{\epsilon}},h_{\boldsymbol{\eta}},\boldsymbol{x}_c}(\boldsymbol{x},\boldsymbol{x}_k)\,\mathrm{d}\sigma_d(\boldsymbol{x}) = 2\hat{p}_1(\hat{\rho})\mathbb{E}_2(p(\boldsymbol{y},\hat{\rho})).$$

PROOF. The argument here is similar to the one used for Lemma 8. Take the limit over $\boldsymbol{\epsilon}$ inside the integral and change the order of integration to yield

$$\widetilde{\lim}_{\boldsymbol{\eta}\to 0}2\hat{p}_1(\hat{\rho})\int_{\mathbb{S}^d}\frac{1}{N}\sum_{k=0}^{N-1}h_{\boldsymbol{\eta}}(\langle\boldsymbol{z},\boldsymbol{x}_c\rangle)\mathbf{1}_{C(\boldsymbol{z},\hat{\rho})}(\boldsymbol{x}_k)\,\mathrm{d}\sigma_d(\boldsymbol{z}).$$

Substituting the decomposition $\boldsymbol{z} = t\boldsymbol{x}_c + \sqrt{1-t^2}\boldsymbol{z}^*$ produces

$$\begin{aligned}
&\widetilde{\lim}_{\boldsymbol{\eta}\to 0}2\hat{p}_1(\hat{\rho})\int_{-1}^{1}\frac{\omega_{d-1}}{\omega_d}(1-t^2)^{d/2-1}h_{\boldsymbol{\eta}}(t)\int_{\mathbb{S}^{d-1}}\frac{1}{N}\sum_{k=0}^{N-1}\mathbf{1}_{C(t\boldsymbol{x}_c+\sqrt{1-t^2}\boldsymbol{z}^*,\hat{\rho})}(\boldsymbol{x}_k)\,\mathrm{d}\sigma_{d-1}(\boldsymbol{z}^*)\,\mathrm{d}t\\
&= 2\hat{p}_1(\hat{\rho})\int_{\mathbb{S}^{d-1}}\frac{1}{N}\sum_{k=0}^{N-1}\mathbf{1}_{C(\tilde{\rho}\boldsymbol{x}_c+\sqrt{1-\tilde{\rho}^2}\boldsymbol{z}^*,\hat{\rho})}(\boldsymbol{x}_k)\,\mathrm{d}\sigma_{d-1}(\boldsymbol{z}^*)\\
&= 2\hat{p}_1(\hat{\rho})\mathbb{E}(p(\boldsymbol{y},\hat{\rho}))
\end{aligned}$$

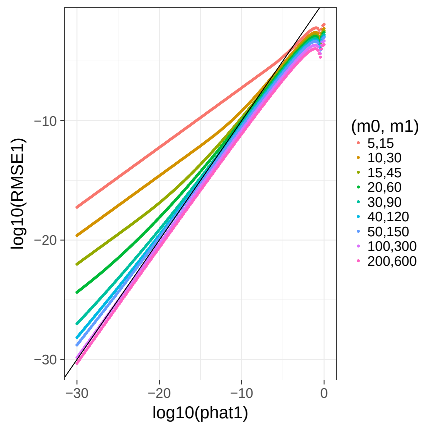
for $\boldsymbol{y}$ under reference distribution 2. □

Proof of Theorem 7

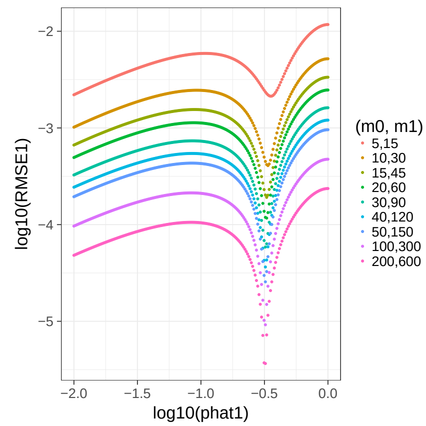
PROOF. The proof follows from using Lemmas 6 to 9 and Theorem 6. □

12. Cases with $\boldsymbol{m_0 \neq m_1}$. Section 7 of the main paper computed our reference moments in some balanced sampling cases, that is cases with $m_0 = m_1$. Here we show results for some unbalanced cases with $m_1 \neq m_0$. There does not appear to be an important difference between balanced and unbalanced cases. The main difference is that given $n = m_0 + m_1$, the balanced cases have smaller values of the granularity limit $1/N$ where $N = \binom{n}{m_0}$.

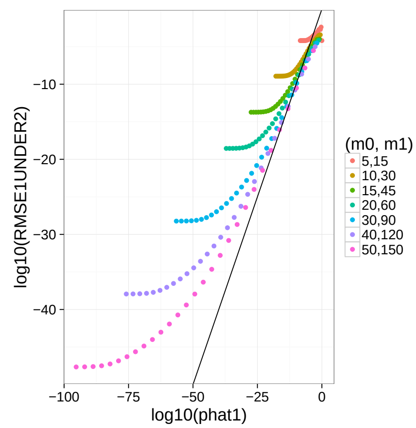
The first illustration has $m_1 = 3m_0$. Figures 8a through 8d are the counterparts to Figures 4a through 4d in Section 7. There are a few minor differences. For $\hat{p}_1$, the 'dip' has in Figure 8b moved from about 10^{-1} to about $10^{-0.5}$, but just as in Figure 4b, the dip is narrow and not located in the important region of smaller p-values. Figure 8c is qualitatively similar to Figure 4c. We see that with the smaller value of m_0, the granularity limit is a larger value.



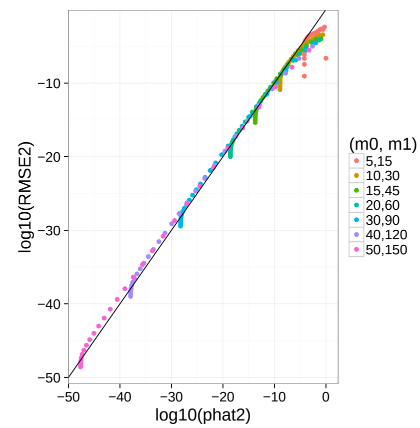
(a) $\mathrm{RMSE}_1(\hat{p}_1)$.



(b) $\mathrm{RMSE}_1(\hat{p}_1)$ zoomed.



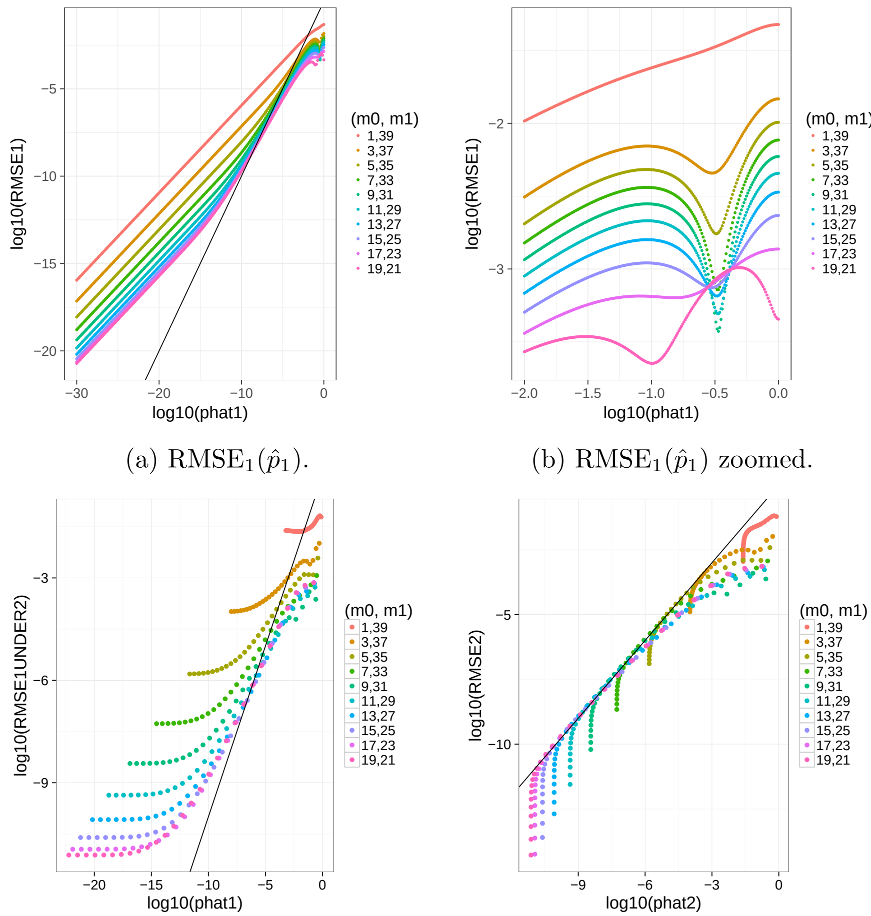
(c) $\mathrm{RMSE}_2(\hat{p}_1)$.



(d) $\mathrm{RMSE}_2(\hat{p}_2)$.

Fig 8: RMSEs for $\hat{p}_1$ and $\hat{p}_2$ under reference distributions 1 and 2. The x-axis shows the estimate $\hat{p}$ as ρ varies from 1 to 0. Here $m_0 = 3m_1$.

For a second illustration we take $m_0 + m_1 = 40$ with $m_0 = 1, 3, 5, \cdots, 19$. The results are in Figures 9a through 9d. The p-values in these figures do not get as small as those in the other figures because the total sample size does not reach the hundreds and because some of the m_0 values are quite small. The asymptotic granularity limits are higher, the dip looks different but is still not in the interesting range, and while the RMSE for $\hat{p}_2$ looks to be further below the 45 degree line than it was for balanced data, that effect is mostly because the range of RMSE values is narrower in this case.



(a) $\text{RMSE}_1(\hat{p}_1)$.

(b) $\text{RMSE}_1(\hat{p}_1)$ zoomed.

(c) $\text{RMSE}_2(\hat{p}_1)$.

(d) $\text{RMSE}_2(\hat{p}_2)$.

Fig 9: RMSEs for $\hat{p}_1$ and $\hat{p}_2$ under reference distributions 1 and 2. The x-axis shows the estimate $\hat{p}$ as ρ varies from 1 to 0. Here $m_0 + m_1 = 40$.

13. Computational cost. Figure 10 shows some empirical running times for our algorithm to compute both $\hat{p}_2 = \mathbb{E}_2(p(\boldsymbol{y}, \hat{\rho}))$ and $\text{Var}_2(p(\boldsymbol{y}, \hat{\rho}))$ in the balanced sample size case. The value of $m = m_0 = m_1$ ranges from 5 to 100 in steps of 5. The running time in seconds is on the vertical axis. These computations were done on an iMac with a 3.2GHz Intel Core i5 processor and 16 Gb of memory. The reference line came from plain least squares regression on the log-log scale. The slope of that regression line is 2.96. Over this range of sample sizes, the computational cost is dominated by the cost of doing $O(m^3)$ integrals and the $O(m^4)$ cost of computing all

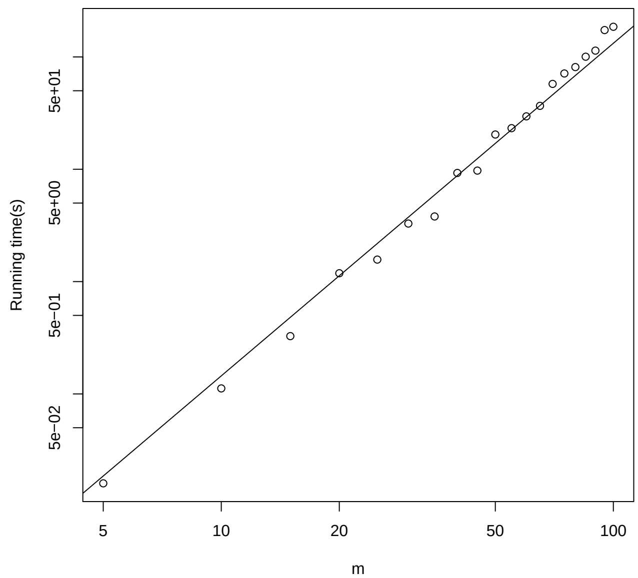


Fig 10: Running time in seconds versus problem size m.

the coefficients $c(r_1, r_2, r_3)$ is not evident.

To have reasonable power to obtain a p-value below ϵ by permutation sampling requires on the order of $1/\epsilon$ permutations, each requiring $O(n)$ computation to generate and $O(n)$ computation to evaluate the inner product. The cost to compute the standard errors in our method is dominated by a cost proportional to $\underline{m}^3$. Assuming that rate applies over the range of important problem sizes, our proposal is advantageous when $\underline{m}^3 = o(n/\epsilon)$. Supposing that m_0 and m_1 are comparable, our advantage holds when $\underline{m}^2 = o(1/\epsilon)$. If only the estimate and not the standard error is required, then our $\hat{p}_2$ and $\hat{p}_3$ cost $O(\underline{m})$ once the $\hat{\rho}$ (cost $O(n)$) has been computed. Then the total cost is $O(n)$ compared to the much larger cost $O(n/\epsilon)$ for sampling. When many tests are being conducted the full $O(\underline{m}^3)$ cost to get a reference variance might only be needed for a handful of the apparently most significant ones.

14. Parkinson's data. We used microarray data from the three cited experiments on Parkinson's disease and gene set lists from the Broad Institute. Our data came from online resources that are subject to change. The gene sets' definitions are continually being updated. The gene expression data sets for Parkinson's disease are also subject to change as new subjects are added. The URLs that we accessed and the data that we have used are available from statweb.stanford.edu/~owen/data/stolarsky. That link also has R code to compute gene set test statistics on these data.

References.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical society*, 68(3):337–404.

Brauchart, J. and Dick, J. (2013). A simple proof of Stolarsky's invariance principle. *Proceedings of the American Mathematical Society*, 141(6):2085–2096.

Stanford University

LinkedIn

University of Pennsylvania

Stanford University