Submitted to the Annals of Statistics

SUPPLEMENTARY MATERIAL FOR 'OPTIMAL NONPARAMETRIC TESTING OF MISSING COMPLETELY AT RANDOM, AND ITS CONNECTIONS TO COMPATIBILITY'

BY THOMAS B. BERRETT[1,*] AND RICHARD J. SAMWORTH[2,†]

[1]*Department of Statistics, University of Warwick, Coventry, CV4 7AL, United Kingdom*
tom.berrett@warwick.ac.uk

[2]*Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, CB3 0WB, United Kingdom*
r.samworth@statslab.cam.ac.uk

This is the supplementary material for Berrett and Samworth (2023).

S1. Proofs and auxiliary results.

PROOF OF THEOREM 2. We apply the idea of *Alexandroff (one-point) compactification* (Alexandroff, 1924). Specifically, writing $\mathcal{J} := \{j \in [d] : \mathcal{X}_j \text{ is not compact}\}$, for each $j \in \mathcal{J}$, we can construct a one-point enlarged space $\mathcal{X}_j^* := \mathcal{X}_j \cup \{\infty_j\}$ (where $\infty_j \notin \mathcal{X}_j$), and take as a topology on $\mathcal{X}_j^*$ all open subsets of $\mathcal{X}_j$ together with all sets of the form $(\mathcal{X}_j \setminus K) \cup \{\infty_j\}$, where K is compact in $\mathcal{X}_j$. With this topology, $\mathcal{X}_j^*$ is a compact, Hausdorff space (Folland, 1999, Proposition 4.36). We also set $\mathcal{X}_j^* := \mathcal{X}_j$ for $j \in [d] \setminus \mathcal{J}$. We can extend each probability measure P_S to a Borel probability measure P_S^* on $\mathcal{X}_S^* := \prod_{j \in S} \mathcal{X}_j^*$ (equipped with the product topology) by setting $P_S^*(B) := P_S(B \cap \mathcal{X}_S)$ for all Borel subsets B of $\mathcal{X}_S^*$.

It is convenient in the first part of this proof to emphasise the underlying spaces by writing, e.g., $\mathcal{G}_{\mathbb{S}}^+(\mathcal{X}_{\mathbb{S}})$, $R_{\mathcal{X}_{\mathbb{S}}}(P_{\mathbb{S}}, f_{\mathbb{S}})$ and $R_{\mathcal{X}_{\mathbb{S}}}(P_{\mathbb{S}})$ in place of $\mathcal{G}_{\mathbb{S}}^+$, $R(P_{\mathbb{S}}, f_{\mathbb{S}})$ and $R(P_{\mathbb{S}})$ respectively. Suppose that $f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+(\mathcal{X}_{\mathbb{S}})$ satisfies $f_S \le |\mathbb{S}| - 1$ for all $S \in \mathbb{S}$. We extend each f_S to a function f_S^* on $\mathcal{X}_S^*$ by defining

$$f_S^*(x_S^*) := \begin{cases} f_S(x_S^*) \text{ if } x_j^* \in \mathcal{X}_j \text{ for all } j \in S \\ |\mathbb{S}| - 1 \text{ otherwise.} \end{cases}$$

To see that f_S^* is upper semi-continuous, first suppose that $x_S^* \in \mathcal{X}_S$ and $y > f_S^*(x_S^*) = f_S(x_S^*)$. Since f_S is upper semi-continuous and all sets that are open in $\mathcal{X}_S$ are open in $\mathcal{X}_S^*$, there exists a neighbourhood $U \subseteq \mathcal{X}_S^*$ of x_S^* such that $f_S^*(x_S) < y$ for all $x_S \in U$. On the other hand, if $x_S^* \in \mathcal{X}_S^* \setminus \mathcal{X}_S$ and $y > f_S^*(x_S^*) = |\mathbb{S}| - 1$, then we can take the neighbourhood $U = \mathcal{X}_S^*$ to see that $f_S^*(x_S) < y$ for all $x_S \in U$. This establishes that f_S^* is indeed upper semi-continuous. Writing $\mathcal{X}^* := \prod_{j \in [d]} \mathcal{X}_j^*$, we also have that

$$\inf_{x^* \in \mathcal{X}^*} \sum_{S \in \mathbb{S}} f_S^*(x_S^*) \ge \min\biggl\{0, \inf_{x \in \mathcal{X}} \sum_{S \in \mathbb{S}} f_S(x_S)\biggr\} = 0,$$

so $f_{\mathbb{S}}^* \in \mathcal{G}_{\mathbb{S}}^+(\mathcal{X}_{\mathbb{S}}^*)$. Moreover,

$$R_{\mathcal{X}_{\mathbb{S}}}(P_{\mathbb{S}}, f_{\mathbb{S}}) = R_{\mathcal{X}_{\mathbb{S}}^*}(P_{\mathbb{S}}^*, f_{\mathbb{S}}^*). \tag{S1}$$

*Research supported by Engineering and Physical Sciences Reseach Council (EPSRC) New Investigator Award EP/W016117/1.

†Research supported by EPSRC Programme grant EP/N031938/1, EPSRC Fellowship EP/P031447/1 and European Research Council Advanced grant 101019498.

In the other direction, given any $f_{\mathbb{S}}^* \in \mathcal{G}_{\mathbb{S}}^+(\mathcal{X}_{\mathbb{S}}^*)$, we can define $f_{\mathbb{S}} = (f_S : S \in \mathbb{S})$ on $\mathcal{X}_{\mathbb{S}}$ by defining each f_S to be the restriction of f_S^* to $\mathcal{X}_S$. Then, for each $t \in \mathbb{R}$,

$$(f_S)^{-1}\big([t,\infty)\big) = (f_S^*)^{-1}\big([t,\infty)\big) \cap \mathcal{X}_S,$$

so $(f_S)^{-1}\big([t,\infty)\big)$ is a closed subset of $\mathcal{X}_S$ and f_S is upper semi-continuous. Moreover,

$$\inf_{x \in \mathcal{X}} \sum_{S \in \mathbb{S}} f_S(x_S) \geq \inf_{x^* \in \mathcal{X}^*} \sum_{S \in \mathbb{S}} f_S^*(x_S^*) \geq 0,$$

so $f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+(\mathcal{X}_{\mathbb{S}})$. Again, the equality (S1) holds. We deduce that

$$\begin{aligned} R_{\mathcal{X}_{\mathbb{S}}}(P_{\mathbb{S}}) &= \sup\big\{R_{\mathcal{X}_{\mathbb{S}}}(P_{\mathbb{S}}, f_{\mathbb{S}}) : f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+(\mathcal{X}_{\mathbb{S}})\big\} \\ &= \sup\big\{R_{\mathcal{X}_{\mathbb{S}}^*}(P_{\mathbb{S}}^*, f_{\mathbb{S}}^*) : f_{\mathbb{S}}^* \in \mathcal{G}_{\mathbb{S}}^+(\mathcal{X}_{\mathbb{S}}^*)\big\} = R_{\mathcal{X}_{\mathbb{S}}^*}(P_{\mathbb{S}}^*). \end{aligned} \tag{S2}$$

Now let $\mathcal{C}_{\mathbb{S}}^+(\mathcal{X}_{\mathbb{S}}^*)$ denote the subset of continuous functions in $\mathcal{G}_{\mathbb{S}}^+(\mathcal{X}_{\mathbb{S}}^*)$. Since compact Hausdorff spaces are completely regular, by Kellerer (1984, Proposition 1.33 and an inspection of the proof of Proposition 3.13), we have

$$R_{\mathcal{X}_{\mathbb{S}}}(P_{\mathbb{S}}) = R_{\mathcal{X}_{\mathbb{S}}^*}(P_{\mathbb{S}}^*) = \sup\big\{R_{\mathcal{X}_{\mathbb{S}}^*}(P_{\mathbb{S}}^*, f_{\mathbb{S}}^*) : f_{\mathbb{S}}^* \in \mathcal{C}_{\mathbb{S}}^+(\mathcal{X}_{\mathbb{S}}^*)\big\}.$$

Having established that $R_{\mathcal{X}_{\mathbb{S}}}(P_{\mathbb{S}})$ may be computed as a supremum over functions defined on compact spaces, we now consider the implications for the dual representation of the one-point compactification. Suppose that $\epsilon \in [0,1]$ is such that $P_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0(\mathcal{X}_{\mathbb{S}}) + \epsilon\mathcal{P}_{\mathbb{S}}(\mathcal{X}_{\mathbb{S}})$. Then $P_{\mathbb{S}} = (1-\epsilon)Q_{\mathbb{S}} + \epsilon T_{\mathbb{S}}$, where $Q_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^0(\mathcal{X}_{\mathbb{S}})$ and $T_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}(\mathcal{X}_{\mathbb{S}})$. For each $S \in \mathbb{S}$, we define probability measures Q_S^*, T_S^* on $\mathcal{X}_S^*$ by $Q_S^*(B) := Q_S(B \cap \mathcal{X}_S)$ and $T_S^*(B) := T_S(B \cap \mathcal{X}_S)$ for all Borel subsets B of $\mathcal{X}_S^*$. Then $Q_{\mathbb{S}}^* \in \mathcal{P}_{\mathbb{S}}^0(\mathcal{X}_{\mathbb{S}}^*)$, because $R_{\mathcal{X}_{\mathbb{S}}^*}(Q_{\mathbb{S}}^*) = R_{\mathcal{X}_{\mathbb{S}}}(Q_{\mathbb{S}}) = 0$ from (S2) and the fact that $Q_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^0(\mathcal{X}_{\mathbb{S}})$. Hence $P_{\mathbb{S}}^* = (1-\epsilon)Q_{\mathbb{S}}^* + \epsilon T_{\mathbb{S}}^* \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0(\mathcal{X}_{\mathbb{S}}^*) + \epsilon\mathcal{P}_{\mathbb{S}}(\mathcal{X}_{\mathbb{S}}^*)$.

Conversely, suppose initially that $\epsilon \in (0,1)$ is such that $P_{\mathbb{S}}^* \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0(\mathcal{X}_{\mathbb{S}}^*) + \epsilon\mathcal{P}_{\mathbb{S}}(\mathcal{X}_{\mathbb{S}}^*)$, so that $P_{\mathbb{S}}^* = (1-\epsilon)Q_{\mathbb{S}}^* + \epsilon T_{\mathbb{S}}^*$, where $Q_{\mathbb{S}}^* \in \mathcal{P}_{\mathbb{S}}^0(\mathcal{X}_{\mathbb{S}}^*)$ and $T_{\mathbb{S}}^* \in \mathcal{P}_{\mathbb{S}}(\mathcal{X}_{\mathbb{S}}^*)$. Observe that we must have $Q_S^*(B) = Q_S^*(B \cap \mathcal{X}_S)$ and $T_S^*(B) = T_S^*(B \cap \mathcal{X}_S)$ for all $S \in \mathbb{S}$ and all Borel subsets $B \subseteq \mathcal{X}_{\mathbb{S}}^*$, because P_S^* does not put any mass outside $\mathcal{X}_S$. Then we can define families of probability measures $Q_{\mathbb{S}} = (Q_S : S \in \mathbb{S})$ and $T_{\mathbb{S}} = (T_S : S \in \mathbb{S})$ by $Q_S(B) := Q_S^*(B)$ and $T_S(B) := T_S^*(B)$ for each $S \in \mathbb{S}$ and each Borel subset B of $\mathcal{X}_{\mathbb{S}}$, and have $P_{\mathbb{S}} = (1-\epsilon)Q_{\mathbb{S}} + \epsilon T_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0(\mathcal{X}_{\mathbb{S}}) + \epsilon\mathcal{P}_{\mathbb{S}}(\mathcal{X}_{\mathbb{S}})$. The boundary cases $\epsilon \in \{0,1\}$ can also be handled similarly, and we deduce that

$$\begin{aligned} &\inf\big\{\epsilon \in [0,1] : P_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0(\mathcal{X}_{\mathbb{S}}) + \epsilon\mathcal{P}_{\mathbb{S}}(\mathcal{X}_{\mathbb{S}})\big\} \\ &\qquad\qquad = \inf\big\{\epsilon \in [0,1] : P_{\mathbb{S}}^* \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0(\mathcal{X}_{\mathbb{S}}^*) + \epsilon\mathcal{P}_{\mathbb{S}}(\mathcal{X}_{\mathbb{S}}^*)\big\}. \end{aligned}$$

The upshot of this argument is that we may assume without loss of generality that each $\mathcal{X}_j$ is a compact Hausdorff space (not just locally compact), so that

$$R(P_{\mathbb{S}}) = \sup\big\{R(P_{\mathbb{S}}, f_{\mathbb{S}}) : f_{\mathbb{S}} \in \mathcal{C}_{\mathbb{S}}^+\big\},$$

where we now have suppressed the dependence of these quantities on $\mathcal{X}_{\mathbb{S}}$. We now seek to apply Isii (1964, Theorem 2.3) to rewrite this expression for $R(P_{\mathbb{S}})$ in its dual form; this will require some further definitions. Let

$$X := \{g_{\mathbb{S}} = (g_S : S \in \mathbb{S}) : g_S : \mathcal{X}_S \to [0,\infty) \text{ is continuous for all } S \in \mathbb{S}\},$$

let Z denote the set of real-valued, continuous functions on $\mathcal{X}$ endowed with the supremum norm topology, let $\mathcal{C} \subseteq Z$ denote those elements of Z that are non-negative, let $\psi : X \to Z$ be given by $\psi(g_{\mathbb{S}})(x) := (1/|\mathbb{S}|)\sum_{S \in \mathbb{S}} g_S(x_S)$, and let $\phi : X \to \mathbb{R}$ be given by $\phi(g_{\mathbb{S}}) :=$

$-(1/|\mathbb{S}|)\sum_{S\in\mathbb{S}}\int g_S\,dP_S$. Now $\mathcal{C}$ is a convex cone with non-empty interior. Moreover, for any $g \in Z$ we can take $g_{\mathbb{S}} = \|g\|_\infty$ and $g' = \|g\|_\infty - g \in \mathcal{C}$ to see that $\psi(g_{\mathbb{S}}) - g' = \|g\|_\infty - g' = g$, and so $\psi(X) - \mathcal{C} = Z$. This shows that Assumption A of Isii (1964) holds. Since X is a convex cone and ϕ and ψ are linear we see that the conditions of Isii (1964, Theorem 2.3) are satisfied. Now, $\mathcal{X}$ is compact by Tychanov's theorem (e.g. Folland, 1999, Theorem 4.42) (which is equivalent to the axiom of choice), so by a version of the Riesz representation theorem (e.g. Folland, 1999, Theorem 7.2), the set of non-negative elements of the continuous dual Z^* of Z is the set of Radon measures on $\mathcal{X}$, denoted $\mathcal{M}_+(\mathcal{X})$. Thus, writing μ^S for the marginal measure on $\mathcal{X}_S$ of $\mu \in \mathcal{M}_+(\mathcal{X})$, we have

$$\begin{aligned} R(P_{\mathbb{S}}) &= 1 + \sup\{\phi(g_{\mathbb{S}}) : g_{\mathbb{S}} \in X, \psi(g_{\mathbb{S}}) - 1 \geq 0\} \\ &= 1 + \inf\bigl\{z^*(-1) : z^* \in Z^*, z^* \geq 0, z^*\bigl(\psi(g_{\mathbb{S}})\bigr) + \phi(g_{\mathbb{S}}) \leq 0 \text{ for all } g_{\mathbb{S}} \in X\bigr\} \\ &= 1 + \inf\biggl\{-\mu(\mathcal{X}) : \mu \in \mathcal{M}_+(\mathcal{X}), \int_{\mathcal{X}} \biggl(\sum_{S\in\mathbb{S}} g_S\biggr) d\mu \leq \sum_{S\in\mathbb{S}} \int_{\mathcal{X}_S} g_S\,dP_S \text{ for all } g_{\mathbb{S}} \in X\biggr\} \end{aligned}$$

(S3)
$$= 1 - \sup\biggl\{\mu(\mathcal{X}) : \mu \in \mathcal{M}_+(\mathcal{X}), \int_{\mathcal{X}_S} g_S\,d\mu^S \leq \int_{\mathcal{X}_S} g_S\,dP_S \text{ for all } S \in \mathbb{S}, g_{\mathbb{S}} \in X\biggr\}.$$

We finally claim that this last display is equal to the claimed form in the statement of the result. Let $\epsilon \in [0,1]$ be such that $P_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0 + \epsilon\mathcal{P}_{\mathbb{S}}$. Then there exists a probability measure μ on $\mathcal{X}$ with marginals $\mu_{\mathbb{S}} := (\mu^S : S \in \mathbb{S})$ for which we can write $P_{\mathbb{S}} = (1-\epsilon)\mu_{\mathbb{S}} + \epsilon Q_{\mathbb{S}}$, where $Q_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$. Since every open set in $\mathcal{X}$ is σ-compact, the probability measure μ is necessarily Radon (Folland, 1999, Theorem 7.8). Now for all $S \in \mathbb{S}$, and $g_{\mathbb{S}} \in X$,

$$(1-\epsilon)\int_{\mathcal{X}_S} g_S\,d\mu^S = \int_{\mathcal{X}_S} g_S\,d(P_S - \epsilon Q_S) \leq \int_{\mathcal{X}_S} g_S\,dP_S$$

so $(1-\epsilon)\mu$ is feasible and we deduce from (S3) that $R(P_{\mathbb{S}}) \leq \epsilon$. Hence $R(P_{\mathbb{S}}) \leq \inf\bigl\{\epsilon \in [0,1] : P_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0 + \epsilon\mathcal{P}_{\mathbb{S}}\bigr\}$. For the bound in the other direction, first suppose that $R(P_{\mathbb{S}}) = 1$. Then, from (S3), the only element μ of $\mathcal{M}_+(\mathcal{X})$ satisfying $\int_{\mathcal{X}_S} g_S\,d\mu^S \leq \int_{\mathcal{X}_S} g_S\,dP_S$ for all $S \in \mathbb{S}, g_{\mathbb{S}} \in X$ is the zero measure on $\mathcal{X}$. If $\epsilon \in [0,1]$ is such that $P_{\mathbb{S}} = (1-\epsilon)Q_{\mathbb{S}} + \epsilon T_{\mathbb{S}}$ with $Q_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^0$ and $T_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$, then for any $S \in \mathbb{S}$ and $g_{\mathbb{S}} \in X$,

$$\int_{\mathcal{X}_S} g_S\,d(1-\epsilon)Q_S \leq \int_{\mathcal{X}_S} g_S\,dP_S.$$

It follows that $(1-\epsilon)Q_{\mathbb{S}} \in \mathcal{M}_+(\mathcal{X})$ must be the zero measure, so $\epsilon = 1$. Hence, when $R(P_{\mathbb{S}}) = 1$, we also have $\inf\bigl\{\epsilon \in [0,1] : P_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0 + \epsilon\mathcal{P}_{\mathbb{S}}\bigr\} = 1$. Now suppose that $R(P_{\mathbb{S}}) < 1$, so by (S3), given $\delta \in \bigl(0, 1 - R(P_{\mathbb{S}})\bigr)$, we can find $\mu \in \mathcal{M}_+(\mathcal{X})$ with marginals $(\mu^S : S \in \mathbb{S})$ that satisfies $\int_{\mathcal{X}_S} g_S\,d\mu^S \leq \int_{\mathcal{X}_S} g_S\,dP_S$ for all $S \in \mathbb{S}, g_{\mathbb{S}} \in X$ and $\mu(\mathcal{X}) = 1 - R(P_{\mathbb{S}}) - \delta$. Writing $\epsilon := 1 - \mu(\mathcal{X}) = R(P_{\mathbb{S}}) + \delta$, let $Q_{\mathbb{S}} := (\mu^S/(1-\epsilon) : S \in \mathbb{S}) \in \mathcal{P}_{\mathbb{S}}^0$, and let $T_{\mathbb{S}} := \epsilon^{-1}\bigl(P_{\mathbb{S}} - (1-\epsilon)Q_{\mathbb{S}}\bigr)$. Then $T_S(\mathcal{X}_S) = 1$ for all $S \in \mathbb{S}$, and for any $S \in \mathbb{S}$ and $g_{\mathbb{S}} \in X$,

$$\int_{\mathcal{X}_S} g_S\,dT_S = \frac{1}{\epsilon}\int_{\mathcal{X}_S} g_S\,d(P_S - \mu^S) \geq 0.$$

Thus T_S is a probability measure on $\mathcal{X}_S$ for all $S \in \mathbb{S}$, so $T_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$ and $P_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0 + \epsilon\mathcal{P}_{\mathbb{S}}$. Since $\delta \in \bigl(0, 1 - R(P_{\mathbb{S}})\bigr)$ was arbitrary, we deduce that $\inf\bigl\{\epsilon \in [0,1] : P_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0 + \epsilon\mathcal{P}_{\mathbb{S}}\bigr\} \leq R(P_{\mathbb{S}})$. This completes the proof. □

PROOF OF PROPOSITION 3. If $f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+}$, then $\min(f_{\mathbb{S}}, |\mathbb{S}|-1) \in \mathcal{G}_{\mathbb{S}}^{+}$, because if this were not the case, then there would exist $x^0 = (x_S^0 : S \in \mathbb{S}) \in \mathcal{X}$ and $S_0 \in \mathbb{S}$ with $f_{S_0}(x_{S_0}^0) > |\mathbb{S}|-1$ such that

$$\sum_{S \in \mathbb{S}} \min\{f_S(x_S^0), |\mathbb{S}|-1\} < 0.$$

But, since $f_{\mathbb{S}} \geq -1$, we would then have

$$\sum_{S \in \mathbb{S}} \min\{f_S(x_S^0), |\mathbb{S}|-1\} > |\mathbb{S}|-1 + \sum_{S \in \mathbb{S}: S \neq S_0} f_S(x_S^0) \geq 0,$$

a contradiction. Since $R\bigl(P_{\mathbb{S}}, \min(f_{\mathbb{S}}, |\mathbb{S}|-1)\bigr) \geq R(P_{\mathbb{S}}, f_{\mathbb{S}})$, it follows that, in seeking a maximiser of $R(P_{\mathbb{S}}, \cdot)$, we may restrict our optimisation to $\bigl\{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+} : f_{\mathbb{S}} \leq |\mathbb{S}|-1\bigr\}$.

Writing $\mathcal{G}_S^{**}$ for the set of real-valued, measurable functions on $\mathcal{X}_S$, we therefore have

$$\begin{aligned}
|R(P_{\mathbb{S}}) - R(Q_{\mathbb{S}})| &\leq \sup_{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+} : f_{\mathbb{S}} \leq |\mathbb{S}|-1} \bigl|R(P_{\mathbb{S}}, f_{\mathbb{S}}) - R(Q_{\mathbb{S}}, f_{\mathbb{S}})\bigr| \\
&= \frac{1}{|\mathbb{S}|} \sup_{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+} : f_{\mathbb{S}} \leq |\mathbb{S}|-1} \biggl|\sum_{S \in \mathbb{S}} \int_{\mathcal{X}_S} f_S \, d(P_S - Q_S)\biggr| \\
&\leq \frac{1}{|\mathbb{S}|} \sum_{S \in \mathbb{S}} \sup_{f_S \in \mathcal{G}_S^{**} : -1 \leq f_S \leq |\mathbb{S}|-1} \biggl|\int_{\mathcal{X}_S} f_S \, d(P_S - Q_S)\biggr| \\
&= \sum_{S \in \mathbb{S}} \sup_{f_S \in \mathcal{G}_S^{**} : -1/2 \leq f_S \leq 1/2} \biggl|\int_{\mathcal{X}_S} f_S \, d(P_S - Q_S)\biggr| = d_{\mathrm{TV}}(P_{\mathbb{S}}, Q_{\mathbb{S}}),
\end{aligned}$$

as required. □

PROOF OF PROPOSITION 4. Our strategy here is to apply results on the concentration properties and the mean of the supremum $R(\widehat{P}_{\mathbb{S}})$ of the empirical process

$$R(\widehat{P}_{\mathbb{S}}, f_{\mathbb{S}}) = -\frac{1}{|\mathbb{S}|} \sum_{S \in \mathbb{S}} \frac{1}{n_S} \sum_{i=1}^{n_S} f_S(X_{S,i}) \tag{S4}$$

over $f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+}$. As in the proof of Proposition 4, we may restrict our optimisation to $\bigl\{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+} : f_{\mathbb{S}} \leq |\mathbb{S}|-1\bigr\}$.

Writing $V := \sum_{S \in \mathbb{S}} n_S^{-1}$, by Boucheron, Lugosi and Massart (2013, Theorem 12.1) — a consequence of the bounded differences (McDiarmid's) inequality — for any collection $P_{\mathbb{S}}$ and $\lambda \in \mathbb{R}$, we have

$$\log \mathbb{E} \exp\bigl(\lambda\bigl\{R(\widehat{P}_{\mathbb{S}}) - \mathbb{E}R(\widehat{P}_{\mathbb{S}})\bigr\}\bigr) \leq \frac{V\lambda^2}{8}.$$

In particular, by the usual sub-Gaussian tail bound,

$$\begin{aligned}
\max\bigl\{\mathbb{P}\bigl(R(\widehat{P}_{\mathbb{S}}) - \mathbb{E}R(\widehat{P}_{\mathbb{S}}) \leq -t\bigr), \mathbb{P}\bigl(R(\widehat{P}_{\mathbb{S}}) - \mathbb{E}R(\widehat{P}_{\mathbb{S}}) \geq t\bigr)\bigr\} &\leq \exp\Bigl(-\frac{2t^2}{V}\Bigr) \\
&= \exp\Bigl(-\frac{2t^2}{\sum_{S \in \mathbb{S}} n_S^{-1}}\Bigr)
\end{aligned} \tag{S5}$$

for all $t \geq 0$. Moreover, by Proposition 3 and two applications of Cauchy–Schwarz,

$$
\begin{aligned}
\bigl|\mathbb{E}R(\widehat{P}_{\mathbb{S}}) - R(P_{\mathbb{S}})\bigr| &\leq \mathbb{E}\bigl|R(\widehat{P}_{\mathbb{S}}) - R(P_{\mathbb{S}})\bigr| \\
&\leq \frac{1}{2}\sum_{S\in\mathbb{S}}\sum_{x_S\in\mathcal{X}_S}\mathbb{E}\bigl|\widehat{P}_S(\{x_S\}) - P_S(\{x_S\})\bigr| \\
&\leq \frac{1}{2}\sum_{S\in\mathbb{S}}\frac{1}{n_S^{1/2}}\sum_{x_S\in\mathcal{X}_S}\bigl[P_S(\{x_S\})\{1-P_S(\{x_S\})\}\bigr]^{1/2} \\
&\leq \frac{1}{2}\sum_{S\in\mathbb{S}}\Bigl(\frac{|\mathcal{X}_S|-1}{n_S}\Bigr)^{1/2}. \qquad \text{(S6)}
\end{aligned}
$$

It follows from (S5) and (S6) that under H_0', i.e. when $R(P_{\mathbb{S}}) = 0$, we have

$$
\mathbb{P}\bigl(R(\widehat{P}_{\mathbb{S}}) \geq C_\alpha\bigr) \leq \mathbb{P}\biggl(R(\widehat{P}_{\mathbb{S}}) - \mathbb{E}R(\widehat{P}_{\mathbb{S}}) \geq \biggl\{\frac{1}{2}\log(1/\alpha)\sum_{S\in\mathbb{S}}\frac{1}{n_S}\biggr\}^{1/2}\biggr) \leq \alpha.
$$

On the other hand, if $R(P_{\mathbb{S}}) \geq C_\alpha + C_\beta$, then from (S5) and (S6) again,

$$
\begin{aligned}
&\mathbb{P}\bigl(R(\widehat{P}_{\mathbb{S}}) \geq C_\alpha\bigr) \\
&\quad\geq \mathbb{P}\biggl(R(\widehat{P}_{\mathbb{S}}) - R(P_{\mathbb{S}}) \geq -\frac{1}{2}\sum_{S\in\mathbb{S}}\Bigl(\frac{|\mathcal{X}_S|-1}{n_S}\Bigr)^{1/2} - \biggl\{\frac{1}{2}\log(1/\beta)\sum_{S\in\mathbb{S}}\frac{1}{n_S}\biggr\}^{1/2}\biggr) \\
&\quad\geq \mathbb{P}\biggl(R(\widehat{P}_{\mathbb{S}}) - \mathbb{E}R(\widehat{P}_{\mathbb{S}}) \geq -\biggl\{\frac{1}{2}\log(1/\beta)\sum_{S\in\mathbb{S}}\frac{1}{n_S}\biggr\}^{1/2}\biggr) \geq 1-\beta,
\end{aligned}
$$

as required. □

PROOF OF PROPOSITION 6. By the same argument given at the start of the proof of Proposition 4, in seeking a maximiser in (2), we may restrict our optimisation to $\bigl\{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+} : f_{\mathbb{S}} \leq |\mathbb{S}| - 1\bigr\}$. But $\bigl[-1, |\mathbb{S}|-1\bigr]^{d_{\mathbb{S}}}$ is a compact subset of $\mathbb{R}^{d_{\mathbb{S}}}$, and we may regard $f_{\mathbb{S}} \mapsto R(P_{\mathbb{S}}, f_{\mathbb{S}})$ as a continuous function on this set, so the supremum in (2) is attained.

By specialising Theorem 2 to the discrete case we see that

$$
R(P_{\mathbb{S}}) = \sup\bigl\{\epsilon \in [0,1] : P_{\mathbb{S}} = \epsilon Q_{\mathbb{S}} + (1-\epsilon)T_{\mathbb{S}}, Q_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^0, T_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}\bigr\}.
$$

When $R(P_{\mathbb{S}}) = 0$ we can trivially attain the supremum by taking $Q_{\mathbb{S}} = P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^0$, since we already know that $R(P_{\mathbb{S}}) = 0$ if and only if $\mathcal{P}_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^0$. Supposing that $R(P_{\mathbb{S}}) > 0$, for each $m \geq 1/R(P_{\mathbb{S}})$ we can find $Q_{\mathbb{S}}^{(m)} \in \mathcal{P}_{\mathbb{S}}^0$, $T_{\mathbb{S}}^{(m)} \in \mathcal{P}_{\mathbb{S}}$, and $\epsilon^{(m)} \in [R(P_{\mathbb{S}}), R(P_{\mathbb{S}}) - 1/m]$ such that $P_{\mathbb{S}} = \epsilon^{(m)}Q_{\mathbb{S}}^{(m)} + (1-\epsilon^{(m)})T_{\mathbb{S}}^{(m)}$. There exists a subsequence $(m_k)_{k\in\mathbb{N}}$, $Q_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^0$, and $T_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$ such that $Q_{\mathbb{S}}^{(m_k)} \to Q_{\mathbb{S}}$ and $T_{\mathbb{S}}^{(m_k)} \to T_{\mathbb{S}}$ as $k \to \infty$. We see that we must have

$$
P_{\mathbb{S}} = R(P_{\mathbb{S}})Q_{\mathbb{S}} + \{1 - R(P_{\mathbb{S}})\}T_{\mathbb{S}},
$$

so that the supremum in (2) is indeed attained.

We now turn to the second part of the result. From Theorem 2 we know that for any $\epsilon > 0$ we have $R(P_{\mathbb{S}}) \leq \epsilon$ if and only if $P_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0 + \epsilon\mathcal{P}_{\mathbb{S}}$. Now suppose that $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$ satisfies $R(P_{\mathbb{S}}) \leq \epsilon$. Then there exist $Q_{\mathbb{S}}^0 \in \mathcal{P}_{\mathbb{S}}^0$ and $Q_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$ such that $P_{\mathbb{S}} = (1-\epsilon)Q_{\mathbb{S}}^0 + \epsilon Q_{\mathbb{S}}$. Since $\mathcal{P}_{\mathbb{S}}^0 \subseteq \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$, it follows that if $S_1, S_2 \in \mathbb{S}$ have $S_1 \cap S_2 \neq \emptyset$, then

$$
Q_{S_1}^{S_1\cap S_2} = \frac{1}{\epsilon}\bigl\{P_{S_1}^{S_1\cap S_2} - (1-\epsilon)Q_{S_1}^{0,S_1\cap S_2}\bigr\} = \frac{1}{\epsilon}\bigl\{P_{S_2}^{S_1\cap S_2} - (1-\epsilon)Q_{S_2}^{0,S_1\cap S_2}\bigr\} = Q_{S_2}^{S_1\cap S_2};
$$

in other words, $Q_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$. Thus, if $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$, then $R(P_{\mathbb{S}}) \leq \epsilon$ if and only if $P_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^{0} + \epsilon\mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$, which holds if and only if

$$P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{0,*} + \epsilon\mathcal{P}_{\mathbb{S}}^{\mathrm{cons},**} = \epsilon(\mathcal{P}_{\mathbb{S}}^{0,*} + \mathcal{P}_{\mathbb{S}}^{\mathrm{cons},**}) =: \mathcal{P}_{\mathbb{S}}^{\epsilon,*}. \tag{S7}$$

Now $\mathcal{P}_{\mathbb{S}}^{1,*}$ is a convex polyhedral set, so there exist $B \in \mathbb{R}^{F \times \mathcal{X}_{\mathbb{S}}}$ and $b \in \mathbb{R}^{F}$ such that

$$\mathcal{P}_{\mathbb{S}}^{\epsilon,*} = \bigl\{p_{\mathbb{S}} \equiv P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons},*} : Bp_{\mathbb{S}} \geq -\epsilon b\bigr\},$$

where the equivalence here indicates that $p_{\mathbb{S}}$ is the probability mass sequence corresponding to $P_{\mathbb{S}}$. Since $0_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\epsilon,*}$, we must have $b \in [0,\infty)^F$ and, by rescaling the rows of B if necessary, we may assume that $b \in \{0,1\}^F$. We may therefore partition $B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$, where $B_1 \in \mathbb{R}^{(F-m)\times\mathcal{X}_{\mathbb{S}}}$ and $B_2 \in \mathbb{R}^{m\times\mathcal{X}_{\mathbb{S}}}$ are such that

$$\mathcal{P}_{\mathbb{S}}^{\epsilon,*} = \bigl\{p_{\mathbb{S}} \equiv P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons},*} : B_1 p_{\mathbb{S}} \geq -\epsilon, B_2 p_{\mathbb{S}} \geq 0\bigr\}. \tag{S8}$$

In fact, however, we claim that $m = 0$, so that $b = 1_F$. To see this, note first that $(\mathcal{P}_{\mathbb{S}}^{\epsilon,*})_{\epsilon\geq 0}$ is an increasing family, by (S8). Moreover, if $\lambda \geq 0$ and $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$, then $\lambda \cdot P_{\mathbb{S}} \in \lambda\mathcal{P}_{\mathbb{S}}^{\mathrm{cons},**} \subseteq \lambda(\mathcal{P}_{\mathbb{S}}^{0,*} + \mathcal{P}_{\mathbb{S}}^{\mathrm{cons},**}) = \mathcal{P}_{\mathbb{S}}^{\lambda,*}$, and hence $\bigcup_{\epsilon\geq 0}\mathcal{P}_{\mathbb{S}}^{\epsilon,*} = \mathcal{P}_{\mathbb{S}}^{\mathrm{cons},*}$. But

$$\bigcup_{\epsilon\geq 0}\mathcal{P}_{\mathbb{S}}^{\epsilon,*} = \bigl\{p_{\mathbb{S}} \equiv P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons},*} : B_2 p_{\mathbb{S}} \geq 0\bigr\},$$

and we conclude that $m = 0$, as required. Therefore, by (S7), when $p_{\mathbb{S}} \equiv P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$, we have

$$R(P_{\mathbb{S}}) = \inf\{\epsilon > 0 : P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\epsilon,*}\} = \|Bp_{\mathbb{S}}\|_\infty. \tag{S9}$$

We now argue that $f_{\mathbb{S}}^{(1)}, \ldots, f_{\mathbb{S}}^{(F)}$ can be taken to be scalar multiples of the rows of B. We may regard $\mathcal{P}_{\mathbb{S}}^{\mathrm{cons},*}$ as a convex cone in $[0,\infty)^{\mathcal{X}_{\mathbb{S}}}$; this cone is not full-dimensional (due to the consistency constraints), but if instead we regard it as a subset of its affine hull, then we will be able to express it uniquely as an intersection of halfspaces. To see this, note that the consistency constraints are linear, so there exist $d_0 \leq |\mathcal{X}_{\mathbb{S}}|$ and $U \in \mathbb{R}^{\mathcal{X}_{\mathbb{S}}\times d_0}$ of full column rank such that

$$\mathcal{P}_{\mathbb{S}}^{\mathrm{cons},*} = \{Uy : Uy \geq 0, y \in \mathbb{R}^{d_0}\}.$$

Writing $f_{\mathbb{S}}^{(1)}, \ldots, f_{\mathbb{S}}^{(M)}$ for the extreme points of $\{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+} : f_{\mathbb{S}} \leq |\mathbb{S}| - 1\}$, we have

$$\begin{aligned} \mathcal{Y}^{1,*} := \{y \in \mathbb{R}^{d_0} : Uy \in \mathcal{P}_{\mathbb{S}}^{1,*}\} &= \{y \in \mathbb{R}^{d_0} : Uy \geq 0, BUy \geq -1\} \\ &= \Bigl\{y \in \mathbb{R}^{d_0} : Uy \geq 0, \min_{\ell\in[M]}(f_{\mathbb{S}}^{(\ell)})^T Uy \geq -|\mathbb{S}|\Bigr\}. \end{aligned}$$

Since $\mathcal{Y}^{1,*}$ is a full-dimensional, convex subset of $\mathbb{R}^{d_0}$, the uniqueness of halfspace representations means that by relabelling if necessary, we may assume that each row of BU is $(f_{\mathbb{S}}^{(\ell)})^T U/|\mathbb{S}|$ for some $\ell \in [F]$. Hence $\mathcal{Y}^{1,*} = \bigl\{y \in \mathbb{R}^{d_0} : Uy \geq 0, \min_{\ell\in[F]}(f_{\mathbb{S}}^{(\ell)})^T Uy \geq -|\mathbb{S}|\bigr\}$, and

$$\mathcal{P}_{\mathbb{S}}^{1,*} = \Bigl\{p_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons},*} : \min_{\ell\in[F]}(f_{\mathbb{S}}^{(\ell)})^T p_{\mathbb{S}} \geq -|\mathbb{S}|\Bigr\}.$$

It therefore follows from (S9) that, when $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$, we have

$$R(P_{\mathbb{S}}) = \max_{\ell\in[F]} R(P_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell)})_{+}. \tag{S10}$$

Having characterised the incompatibility index for consistent distributions, we finally prove the given bounds on this index in the general case. To see the lower bound, let $S_1, S_2 \in \mathbb{S}$ be such that $S_1 \cap S_2 \neq \emptyset$, and let $E \subseteq \mathcal{X}_{S_1 \cap S_2}$. Define $f_{\mathbb{S}}^{S_1,S_2,E} = (f_S^{S_1,S_2,E} : S \in \mathbb{S}) \in \mathcal{G}_{\mathbb{S}}$ by

$$f_S^{S_1,S_2,E}(x_S) := \begin{cases} 1 & \text{if } S = S_1,\, x_{S_1 \cap S_2} \in E \\ -1 & \text{if } S = S_2,\, x_{S_1 \cap S_2} \in E \\ 0 & \text{otherwise.} \end{cases}$$

It is straightforward to check that $f_{\mathbb{S}}^{S_1,S_2,E} \in \mathcal{G}_{\mathbb{S}}^+$: if $x \in \mathcal{X}$ is such that $x_{S_1 \cap S_2} \in E$ then

$$\sum_{S \in \mathbb{S}} f_S^{S_1,S_2,E}(x_S) = f_{S_1}^{S_1,S_2,E}(x_{S_1}) + f_{S_2}^{S_1,S_2,E}(x_{S_2}) = 1 - 1 = 0,$$

and if x is such that $x_{S_1 \cap S_2} \notin E$ then $f_S^{S_1,S_2,E}(x_S) = 0$ for all $S \in \mathbb{S}$. We also have that

$$\begin{aligned} R(P_{\mathbb{S}}, f_{\mathbb{S}}^{S_1,S_2,E}) &= -\frac{1}{|\mathbb{S}|}\biggl\{ \sum_{x_{S_1} \in \mathcal{X}_{S_1} : x_{S_1 \cap S_2} \in E} P_{S_1}(\{x_{S_1}\}) - \sum_{x_{S_2} \in \mathcal{X}_{S_2} : x_{S_1 \cap S_2} \in E} P_{S_2}(\{x_{S_2}\}) \biggr\} \\ &= \frac{1}{|\mathbb{S}|}\bigl\{ P_{S_2}^{S_1 \cap S_2}(E) - P_{S_1}^{S_1 \cap S_2}(E) \bigr\}. \end{aligned}$$

We conclude that

$$\begin{aligned} R(P_{\mathbb{S}}) &\geq \max\biggl\{ \max_{\ell \in [F]} R(P_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell)})_+, \max_{S_1,S_2 \in \mathbb{S}: S_1 \cap S_2 \neq \emptyset} \max_{E \subseteq \mathcal{X}_{S_1 \cap S_2}} R(P_{\mathbb{S}}, f_{\mathbb{S}}^{S_1,S_2,E}) \biggr\} \\ &= \max\biggl\{ \max_{\ell \in [F]} R(P_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell)})_+, \frac{1}{|\mathbb{S}|} \max_{S_1,S_2 \in \mathbb{S}: S_1 \cap S_2 \neq \emptyset} d_{\mathrm{TV}}\bigl(P_{S_1}^{S_1 \cap S_2}, P_{S_2}^{S_1 \cap S_2}\bigr) \biggr\}. \end{aligned}$$

This establishes the lower bound, and we now turn to the upper bound. Given sequences of signed measures $P_{\mathbb{S}}, Q_{\mathbb{S}} \in \{\lambda_1 \mathcal{P}_{\mathbb{S}} - \lambda_2 \mathcal{P}_{\mathbb{S}} : \lambda_1, \lambda_2 \geq 0\}$, we define their total variation distance by

$$d_{\mathrm{TV}}(P_{\mathbb{S}}, Q_{\mathbb{S}}) := \sum_{S \in \mathbb{S}} \sup_{A_S \in \mathcal{A}_S} |P_S(A_S) - Q_S(A_S)|.$$

Now, given any $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$ and $P_{\mathbb{S}}^{\mathrm{cons},*} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons},*}$, we have by (S10) and the fact (quoted at the start of the proof) that all extreme points of $\mathcal{G}_{\mathbb{S}}^+$ take values in $[-1, |\mathbb{S}|-1]^{\mathcal{X}_{\mathbb{S}}}$ that

$$\begin{aligned} R(P_{\mathbb{S}}) &= \frac{1}{|\mathbb{S}|} \sup_{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+} \{-f_{\mathbb{S}}^T(p_{\mathbb{S}} - p_{\mathbb{S}}^{\mathrm{cons},*} + p_{\mathbb{S}}^{\mathrm{cons},*})\} \\ &\leq \frac{1}{|\mathbb{S}|}\biggl[\sup_{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+} \{-f_{\mathbb{S}}^T(p_{\mathbb{S}} - p_{\mathbb{S}}^{\mathrm{cons},*})\} + \sup_{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+} (-f_{\mathbb{S}}^T p_{\mathbb{S}}^{\mathrm{cons},*}) \biggr] \\ &= \frac{1}{|\mathbb{S}|}\biggl[\sup_{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+} \{-f_{\mathbb{S}}^T(p_{\mathbb{S}} - p_{\mathbb{S}}^{\mathrm{cons},*})\} + \max_{\ell \in [F]} \{-(f_{\mathbb{S}}^{(\ell)})^T(p_{\mathbb{S}}^{\mathrm{cons},*} - p_{\mathbb{S}} + p_{\mathbb{S}})\} \biggr] \\ &\leq \frac{1}{|\mathbb{S}|}\biggl[\sup_{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+} \{-f_{\mathbb{S}}^T(p_{\mathbb{S}} - p_{\mathbb{S}}^{\mathrm{cons},*})\} + \sup_{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+} \{-f_{\mathbb{S}}^T(p_{\mathbb{S}}^{\mathrm{cons},*} - p_{\mathbb{S}})\} + \max_{\ell \in [F]} \{-(f_{\mathbb{S}}^{(\ell)})^T p_{\mathbb{S}}\} \biggr] \end{aligned}$$

(S11)

$$\leq 2 d_{\mathrm{TV}}(P_{\mathbb{S}}, P_{\mathbb{S}}^{\mathrm{cons},*}) + \frac{1}{|\mathbb{S}|} \max_{\ell \in [F]} \{-(f_{\mathbb{S}}^{(\ell)})^T p_{\mathbb{S}}\}.$$

We proceed by constructing an element of $\mathcal{P}_{\mathbb{S}}^{\mathrm{cons},*}$ whose total variation distance to $P_{\mathbb{S}}$ can be controlled. For $\omega \in \{0,1\}^{\mathbb{S}}$, write $T_\omega := \cap_{S:\omega_S=1} S$ and $|\omega| := \sum_{S\in\mathbb{S}} \omega_S$. Define $\tilde{p}_{\mathbb{S}} \in \mathbb{R}^{\mathcal{X}_{\mathbb{S}}}$ by

$$\tilde{p}_{S_0}(x_{S_0}) := p_{S_0}(x_{S_0}) + \sum_{\omega\in\{0,1\}^{\mathbb{S}}:\omega_{S_0}=1, T_\omega\neq\emptyset} \frac{\lambda_{|\omega|}|\mathcal{X}_{T_\omega}|}{|\mathcal{X}_{S_0}|} \sum_{S:\omega_S=1} \{p_S^{T_\omega}(x_{T_\omega}) - p_{S_0}^{T_\omega}(x_{T_\omega})\}$$

with $\lambda_{|\omega|} := \frac{(-1)^{|\omega|}}{|\omega|(|\omega|-1)} \mathbb{1}_{\{|\omega|\geq 2\}}$. Although $\tilde{p}_{\mathbb{S}}$ may take negative values, we will see that it satisfies all the linear constraints of consistency. To see this, let $S_1, S_2 \in \mathbb{S}$ be such that $S_1 \cap S_2 \neq \emptyset$ and $x_{S_1\cap S_2} \in \mathcal{X}_{S_1\cap S_2}$, and write $\Omega_{\mathbb{S}}^{a,b} := \{\omega \in \{0,1\}^{\mathbb{S}} : T_\omega \neq \emptyset, \omega_{S_1} = a, \omega_{S_2} = b\}$ for $a, b \in \{0,1\}$. Observe that if $A \subseteq B \subseteq [d]$, then $|\mathcal{X}_B|/|\mathcal{X}_A| = |\mathcal{X}_{B\cap A^c}|$. Thus, in particular, when $\omega \in \Omega_{\mathbb{S}}^{1,0}$ for instance, we have

$$\frac{|\mathcal{X}_{T_\omega}||\mathcal{X}_{S_1\cap S_2^c\cap T_\omega^c}||\mathcal{X}_{S_1\cap S_2}|}{|\mathcal{X}_{T_\omega\cap S_2}||\mathcal{X}_{S_1}|} = \frac{|\mathcal{X}_{T_\omega\cap S_2^c}||\mathcal{X}_{S_1\cap S_2^c\cap T_\omega^c}|}{|\mathcal{X}_{S_1\cap S_2^c}|} = \frac{|\mathcal{X}_{T_\omega\cap S_2^c}|}{|\mathcal{X}_{S_1\cap S_2^c\cap T_\omega}|} = 1.$$

Hence

$$\begin{aligned}
\tilde{p}_{S_1}^{S_1\cap S_2}(x_{S_1\cap S_2}) - \tilde{p}_{S_2}^{S_1\cap S_2}(x_{S_1\cap S_2}) &= \sum_{\substack{x_{S_1}\in\mathcal{X}_{S_1}:\\ (x_{S_1})_{S_1\cap S_2}=x_{S_1\cap S_2}}} \tilde{p}_{S_1}(x_{S_1}) - \sum_{\substack{x_{S_2}\in\mathcal{X}_{S_2}:\\ (x_{S_2})_{S_1\cap S_2}=x_{S_1\cap S_2}}} \tilde{p}_{S_2}(x_{S_2}) \\
&= p_{S_1}^{S_1\cap S_2}(x_{S_1\cap S_2}) - p_{S_2}^{S_1\cap S_2}(x_{S_1\cap S_2}) \\
&\quad + \sum_{\omega\in\Omega_{\mathbb{S}}^{1,1}} \lambda_{|\omega|} \sum_{S:\omega_S=1} \biggl[\frac{|\mathcal{X}_{T_\omega}|}{|\mathcal{X}_{S_1}|}|\mathcal{X}_{S_1\cap S_2^c}|\{p_S^{T_\omega}(x_{T_\omega}) - p_{S_1}^{T_\omega}(x_{T_\omega})\} \\
&\qquad - \frac{|\mathcal{X}_{T_\omega}|}{|\mathcal{X}_{S_2}|}|\mathcal{X}_{S_1^c\cap S_2}|\{p_S^{T_\omega}(x_{T_\omega}) - p_{S_2}^{T_\omega}(x_{T_\omega})\}\biggr] \\
&\quad + \sum_{\omega\in\Omega_{\mathbb{S}}^{1,0}} \lambda_{|\omega|} \sum_{S:\omega_S=1} \frac{|\mathcal{X}_{T_\omega}|}{|\mathcal{X}_{S_1}|}|\mathcal{X}_{S_1\cap S_2^c\cap T_\omega^c}|\{p_S^{T_\omega\cap S_2}(x_{T_\omega\cap S_2}) - p_{S_1}^{T_\omega\cap S_2}(x_{T_\omega\cap S_2})\} \\
&\quad - \sum_{\omega\in\Omega_{\mathbb{S}}^{0,1}} \lambda_{|\omega|} \sum_{S:\omega_S=1} \frac{|\mathcal{X}_{T_\omega}|}{|\mathcal{X}_{S_2}|}|\mathcal{X}_{S_1^c\cap S_2\cap T_\omega^c}|\{p_S^{T_\omega\cap S_1}(x_{T_\omega\cap S_1}) - p_{S_2}^{T_\omega\cap S_1}(x_{T_\omega\cap S_1})\} \\
&= p_{S_1}^{S_1\cap S_2}(x_{S_1\cap S_2}) - p_{S_2}^{S_1\cap S_2}(x_{S_1\cap S_2}) - \sum_{\omega\in\Omega_{\mathbb{S}}^{1,1}} |\omega|\lambda_{|\omega|}\frac{|\mathcal{X}_{T_\omega}|}{|\mathcal{X}_{S_1\cap S_2}|}\{p_{S_1}^{T_\omega}(x_{T_\omega}) - p_{S_2}^{T_\omega}(x_{T_\omega})\} \\
&\quad + \sum_{\omega'\in\Omega_{\mathbb{S}}^{1,1}} \lambda_{|\omega'|-1}\frac{|\mathcal{X}_{T_{\omega'}}|}{|\mathcal{X}_{S_1\cap S_2}|} \sum_{S:\omega'_S=1} (1-\mathbb{1}_{\{S=S_2\}})\{p_S^{T_{\omega'}}(x_{T_{\omega'}}) - p_{S_1}^{T_{\omega'}}(x_{T_{\omega'}})\} \\
&\quad - \sum_{\omega'\in\Omega_{\mathbb{S}}^{1,1}} \lambda_{|\omega'|-1}\frac{|\mathcal{X}_{T_{\omega'}}|}{|\mathcal{X}_{S_1\cap S_2}|} \sum_{S:\omega'_S=1} (1-\mathbb{1}_{\{S=S_1\}})\{p_S^{T_{\omega'}}(x_{T_{\omega'}}) - p_{S_2}^{T_{\omega'}}(x_{T_{\omega'}})\} \\
&= p_{S_1}^{S_1\cap S_2}(x_{S_1\cap S_2}) - p_{S_2}^{S_1\cap S_2}(x_{S_1\cap S_2}) - \sum_{\omega\in\Omega_{\mathbb{S}}^{1,1}} |\omega|\lambda_{|\omega|}\frac{|\mathcal{X}_{T_\omega}|}{|\mathcal{X}_{S_1\cap S_2}|}\{p_{S_1}^{T_\omega}(x_{T_\omega}) - p_{S_2}^{T_\omega}(x_{T_\omega})\} \\
&\quad - \sum_{\omega\in\Omega_{\mathbb{S}}^{1,1}} \lambda_{|\omega|-1}(|\omega|-2)\frac{|\mathcal{X}_{T_\omega}|}{|\mathcal{X}_{S_1\cap S_2}|}\{p_{S_1}^{T_\omega}(x_{T_\omega}) - p_{S_2}^{T_\omega}(x_{T_\omega})\} = 0,
\end{aligned}$$

where the final equality holds because (λ_r) satisfies $\lambda_2 = 1/2$ and $r\lambda_r = -(r-2)\lambda_{r-1}$ for $r \geq 3$. The total negative mass of $\tilde{p}_{\mathbb{S}}$ satisfies

$$\sum_{S_0 \in \mathbb{S}} \sum_{x_{S_0} \in \mathcal{X}_{S_0}} \tilde{p}_{S_0}(x_{S_0})_- \leq d_{\mathrm{TV}}(P_{\mathbb{S}}, \tilde{P}_{\mathbb{S}})$$

$$\leq \sum_{S_0 \in \mathbb{S}} \sum_{\substack{\omega : |\omega| \geq 2, \\ \omega_{S_0} = 1, T_\omega \neq \emptyset}} \frac{|\mathcal{X}_{T_\omega}|}{|\mathcal{X}_{S_0}||\omega|(|\omega|-1)} \sum_{S : \omega_S = 1} \sum_{x_{S_0} \in \mathcal{X}_{S_0}} \bigl[(-1)^{|\omega|}\{p_S^{T_\omega}(x_{T_\omega}) - p_{S_0}^{T_\omega}(x_{T_\omega})\}\bigr]_-$$

$$\leq \sum_{S_0 \in \mathbb{S}} \sum_{\substack{\omega : |\omega| \geq 2, \\ \omega_{S_0} = 1, T_\omega \neq \emptyset}} \frac{1}{|\omega| - 1} \max_{S : \omega_S = 1} d_{\mathrm{TV}}(p_S^{T_\omega}, p_{S_0}^{T_\omega})$$

$$\leq \Biggl(\sum_{\omega : |\omega| \geq 2, T_\omega \neq \emptyset} \frac{|\omega|}{|\omega| - 1} \Biggr) \max_{S, S_0 \in \mathbb{S} : S \cap S_0 \neq \emptyset} d_{\mathrm{TV}}(p_S^{S \cap S_0}, p_{S_0}^{S \cap S_0})$$

(S12)
$$\leq 2^{|\mathbb{S}|+1} \max_{S, S_0 \in \mathbb{S} : S \cap S_0 \neq \emptyset} d_{\mathrm{TV}}(p_S^{S \cap S_0}, p_{S_0}^{S \cap S_0}).$$

Now define $\check{P}_{\mathbb{S}} \in \{\lambda \cdot \mathcal{P}_{\mathbb{S}} : \lambda \geq 0\}$ with mass function $\check{p}_{\mathbb{S}}$ given by

$$\check{p}_{\mathbb{S}} := \tilde{p}_{\mathbb{S}} + \mathbb{A}\Biggl(\sum_{x \in \mathcal{X}} \delta_x \sum_{S \in \mathbb{S}} \frac{\tilde{p}_S(x_S)_-}{|\mathcal{X}_{S^c}|} \Biggr)$$

where $\delta_y \in \{0,1\}^{\mathcal{X}}$ denotes a Dirac point mass on $y \in \mathcal{X}$. We see that this is non-negative by writing

$$\check{p}_S(x_S) = \tilde{p}_S(x_S) + \sum_{y : y_S = x_S} \sum_{T \in \mathbb{S}} \frac{\tilde{p}_T(y_T)_-}{|\mathcal{X}_{T^c}|} \geq \tilde{p}_S(x_S) + \tilde{p}_S(x_S)_- \geq 0.$$

Since $\tilde{p}_{\mathbb{S}}$ satisfies the consistency constraints and $\check{p}_{\mathbb{S}}$ is formed by adding a compatible sequence of marginal measures to it, we have $\check{P}_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons},*}$. Moreover, $\check{p}_{\mathbb{S}} \geq \tilde{p}_{\mathbb{S}}$ and

$$\sum_{S \in \mathbb{S}} \sum_{x_S \in \mathcal{X}_S} \{\check{p}_S(x_S) - \tilde{p}_S(x_S)\} = 1_{\mathcal{X}_{\mathbb{S}}}^T \mathbb{A}\Biggl(\sum_{x \in \mathcal{X}} \delta_x \sum_{S \in \mathbb{S}} \frac{\tilde{p}_S(x_S)_-}{|\mathcal{X}_{S^c}|} \Biggr)$$

$$= |\mathbb{S}| 1_{\mathcal{X}}^T \Biggl(\sum_{x \in \mathcal{X}} \delta_x \sum_{S \in \mathbb{S}} \frac{\tilde{p}_S(x_S)_-}{|\mathcal{X}_{S^c}|} \Biggr)$$

$$= |\mathbb{S}| \sum_{x \in \mathcal{X}} \sum_{S \in \mathbb{S}} \frac{\tilde{p}_S(x_S)_-}{|\mathcal{X}_{S^c}|}$$

$$= |\mathbb{S}| \sum_{S \in \mathbb{S}} \sum_{x_S \in \mathcal{X}_S} \tilde{p}_S(x_S)_-$$

$$\leq |\mathbb{S}| 2^{|\mathbb{S}|+1} \max_{S_1, S_2 \in \mathbb{S} : S_1 \cap S_2 \neq \emptyset} d_{\mathrm{TV}}(p_{S_1}^{S_1 \cap S_2}, p_{S_2}^{S_1 \cap S_2}).$$

From this and (S12), we conclude that

$$d_{\mathrm{TV}}(P_{\mathbb{S}}, \mathcal{P}^{\mathrm{cons},*}) \leq |\mathbb{S}| 2^{|\mathbb{S}|+2} \max_{S_1, S_2 \in \mathbb{S} : S_1 \cap S_2 \neq \emptyset} d_{\mathrm{TV}}(p_{S_1}^{S_1 \cap S_2}, p_{S_2}^{S_1 \cap S_2}),$$

and the result follows. □

PROOF OF THEOREM 7. We prove the result when $F' \geq 1$, and note that if $F' = 0$ then simpler arguments apply. By Proposition 6 and the discussion after (5), we have

$$\begin{aligned}
&\mathbb{P}\big(R(\widehat{P}_{\mathbb{S}}) \geq C'_\alpha\big) \\
&\leq \mathbb{P}\biggl(D_R \max_{\ell\in[F']} R(\widehat{P}_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell),'}) \geq \frac{C'_\alpha}{2}\biggr) + \mathbb{P}\biggl(\max_{S_1,S_2\in\mathbb{S}} d_{\mathrm{TV}}(\widehat{P}_{S_1}^{S_1\cap S_2}, \widehat{P}_{S_2}^{S_1\cap S_2}) \geq \frac{C'_\alpha}{|\mathbb{S}|2^{|\mathbb{S}|+3}}\biggr) \\
&\leq F' \max_{\ell\in[F']} \mathbb{P}\biggl(R(\widehat{P}_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell),'}) \geq \frac{C'_\alpha}{2D_R}\biggr) \\
&\qquad + \frac{|\mathbb{S}|(|\mathbb{S}|-1)}{2} \max_{S_1,S_2\in\mathbb{S}} \mathbb{P}\biggl(d_{\mathrm{TV}}(\widehat{P}_{S_1}^{S_1\cap S_2}, \widehat{P}_{S_2}^{S_1\cap S_2}) \geq \frac{C'_\alpha}{|\mathbb{S}|2^{|\mathbb{S}|+3}}\biggr). \qquad \text{(S13)}
\end{aligned}$$

Observe that when $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^0$, we have for any $f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+$ that $\mathbb{E}R(\widehat{P}_{\mathbb{S}}, f_{\mathbb{S}}) = R(P_{\mathbb{S}}, f_{\mathbb{S}}) \leq 0$. By (S4) and Hoeffding's inequality, whenever $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^0$, we have for any $\ell \in [F']$ that

$$\begin{aligned}
\mathbb{P}\biggl(R(\widehat{P}_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell),'}) \geq \frac{C'_\alpha}{2D_R}\biggr) &\leq \mathbb{P}\biggl(R(\widehat{P}_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell),'}) - \mathbb{E}R(\widehat{P}_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell),'}) \geq \frac{C'_\alpha}{2D_R}\biggr) \\
&\leq |\mathbb{S}| \max_{S\in\mathbb{S}} \mathbb{P}\biggl(-\frac{1}{n_S}\sum_{i=1}^{n_S}\bigl\{f_S^{(\ell),'}(X_{S,i}) - \mathbb{E}f_S^{(\ell),'}(X_{S,i})\bigr\} \geq \frac{C'_\alpha}{2D_R}\biggr) \\
&\leq |\mathbb{S}| \max_{S\in\mathbb{S}} \exp\biggl(-\frac{n_S(C'_\alpha/D_R)^2}{2|\mathbb{S}|^2}\biggr) \leq \frac{\alpha}{2F'}.
\end{aligned}$$

For the second term in (S13), under H_0', for any $S_1, S_2 \in \mathbb{S}$ with $S_1 \neq S_2$ and $S_1 \cap S_2 \neq \emptyset$, we have

$$\begin{aligned}
&\mathbb{P}\biggl(d_{\mathrm{TV}}(\widehat{P}_{S_1}^{S_1\cap S_2}, \widehat{P}_{S_2}^{S_1\cap S_2}) \geq \frac{C'_\alpha}{|\mathbb{S}|2^{|\mathbb{S}|+3}}\biggr) \\
&= \mathbb{P}\biggl(\max_{A\subseteq\mathcal{X}_{S_1\cap S_2}} \bigl|\widehat{P}_{S_1}^{S_1\cap S_2}(A) - P_{S_1}^{S_1\cap S_2}(A) + P_{S_2}^{S_1\cap S_2}(A) - \widehat{P}_{S_2}^{S_1\cap S_2}(A)\bigr| \geq \frac{C'_\alpha}{|\mathbb{S}|2^{|\mathbb{S}|+3}}\biggr) \\
&\leq 2^{|\mathcal{X}_{S_1\cap S_2}|} \max_{A\subseteq\mathcal{X}_{S_1\cap S_2}} \max_{k\in\{1,2\}} \mathbb{P}\biggl(\bigl|\widehat{P}_{S_k}^{S_1\cap S_2}(A) - P_{S_k}^{S_1\cap S_2}(A)\bigr| \geq \frac{C'_\alpha}{|\mathbb{S}|2^{|\mathbb{S}|+4}}\biggr) \\
&\leq 2^{|\mathcal{X}_{S_1\cap S_2}|+1} \exp\biggl(-\frac{(n_{S_1}\wedge n_{S_2})(C'_\alpha)^2}{|\mathbb{S}|^2 2^{2|\mathbb{S}|+7}}\biggr) \leq \frac{\alpha}{|\mathbb{S}|(|\mathbb{S}|-1)},
\end{aligned}$$

where we have used the fact that $\bigl|\widehat{P}_{S_k}^{S_1\cap S_2}(A) - P_{S_k}^{S_1\cap S_2}(A)\bigr| = \bigl|\widehat{P}_{S_k}^{S_1\cap S_2}(A^c) - P_{S_k}^{S_1\cap S_2}(A^c)\bigr|$, and where the penultimate bound follows from Hoeffding's inequality. We have now established that $\mathbb{P}\big(R(\widehat{P}_{\mathbb{S}}) \geq C'_\alpha\big) \leq \alpha$ whenever $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^0$.

We now turn to the final part of Proposition 7. Very similar arguments to those above based on Hoeffding's inequality show that

$$\mathbb{P}\biggl(\max_{\ell\in[F']} R(\widehat{P}_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell),'}) < C'_\alpha\biggr) \leq \beta$$

whenever

$$\max_{\ell\in[F']} R(P_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell),'}) \geq C'_\alpha + |\mathbb{S}|\biggl\{\frac{2\log(F'|\mathbb{S}|/\beta)}{\min_{S\in\mathbb{S}} n_S}\biggr\}^{1/2}.$$

Likewise, for any $S_1, S_2 \in \mathbb{S}$ with $S_1 \cap S_2 \neq \emptyset$,

$$\mathbb{P}\Big(d_{\mathrm{TV}}(\widehat{P}_{S_1}^{S_1\cap S_2}, \widehat{P}_{S_2}^{S_1\cap S_2}) < |\mathbb{S}|C'_\alpha\Big) \leq \beta$$

whenever

$$d_{\mathrm{TV}}(P_{S_1}^{S_1\cap S_2}, P_{S_2}^{S_1\cap S_2}) \geq |\mathbb{S}|C'_\alpha + \left\{\frac{2}{n_{S_1} \wedge n_{S_2}} \log\left(\frac{2^{|\mathcal{X}_{S_1\cap S_2}|+1}}{\beta}\right)\right\}^{1/2}.$$

Now, by Proposition 6, if $R(P_\mathbb{S}) \geq M(C'_\alpha + C'_\beta)$ then we must either have

$$\max_{\ell\in[F']} R(P_\mathbb{S}, f_\mathbb{S}^{(\ell),\prime}) \geq \frac{M}{2D_R}(C'_\alpha + C'_\beta)$$

or

$$\max_{S_1,S_2\in\mathbb{S}} d_{\mathrm{TV}}\big(P_{S_1}^{S_1\cap S_2}, P_{S_2}^{S_1\cap S_2}\big) \geq \frac{M}{2^{|\mathbb{S}|+3}|\mathbb{S}|}(C'_\alpha + C'_\beta).$$

Since

$$C'_\beta \asymp_{|\mathbb{S}|} |\mathbb{S}|D_R\left\{\frac{2\log(F'|\mathbb{S}|/\beta)}{\min_{S\in\mathbb{S}} n_S}\right\}^{1/2} + \max_{\substack{S_1,S_2\in\mathbb{S}:\\ S_1\cap S_2\neq\emptyset}} \left\{\frac{2}{n_{S_1} \wedge n_{S_2}} \log\left(\frac{2^{|\mathcal{X}_{S_1\cap S_2}|+1}}{\beta}\right)\right\}^{1/2},$$

the result follows. □

PROOF OF THEOREM 8. We establish the equality (7) by providing matching upper and lower bounds, first providing the required lower bound on $R(P_\mathbb{S})$. Given $A \subseteq [r]$ and $B \subseteq [s]$, we construct $f_\mathbb{S} \in \mathcal{G}_\mathbb{S}$ as follows. Writing, for example, $f_{ij\bullet} := f_{\{1,2\}}(i,j)$, define

$$f_{ij\bullet} := \begin{cases} 2 & \text{if } (i,j) \in A \times B \\ -1 & \text{if } (i,j) \in A \times B^c \\ -1 & \text{if } (i,j) \in A^c \times B \\ 2 & \text{if } (i,j) \in A^c \times B^c \end{cases}, \quad (f_{i\bullet 1}, f_{i\bullet 2}) := \begin{cases} (-1,2) & \text{if } i \in A \\ (2,-1) & \text{if } i \in A^c \end{cases},$$

and

$$(f_{\bullet j1}, f_{\bullet j2}) := \begin{cases} (-1,2) & \text{if } j \in B \\ (2,-1) & \text{if } j \in B^c \end{cases}.$$

It is straightforward to check that $f_\mathbb{S} \in \mathcal{G}_\mathbb{S}^+$ because, for instance, if $i \in A$ and $j \in B$, then

$$\min(f_{ij\bullet} + f_{\bullet j1} + f_{i\bullet 1}, f_{ij\bullet} + f_{\bullet j2} + f_{i\bullet 2}) = \min(2-1-1, 2+2+2) = 0.$$

Hence

$$\begin{aligned}
3R(P_\mathbb{S}) &\geq 3R(P_\mathbb{S}, f_\mathbb{S}) \\
&= -\sum_{i=1}^r\sum_{j=1}^s p_{ij\bullet}f_{ij\bullet} - \sum_{i=1}^r (p_{i\bullet 1}f_{i\bullet 1} + p_{i\bullet 2}f_{i\bullet 2}) - \sum_{j=1}^s (p_{\bullet j1}f_{\bullet j1} + p_{\bullet j2}f_{\bullet j2}) \\
&= -2(p_{AB\bullet} + p_{A^cB^c\bullet}) + (p_{A^cB\bullet} + p_{AB^c\bullet}) - 2(p_{A^c\bullet 1} + p_{A\bullet 2}) \\
&\qquad + (p_{A\bullet 1} + p_{A^c\bullet 2}) - 2(p_{\bullet B^c 1} + p_{\bullet B2}) + (p_{\bullet B1} + p_{\bullet B^c 2}) \\
&= -2(2p_{AB\bullet} + p_{\bullet\bullet\bullet} - p_{A\bullet\bullet} - p_{\bullet B\bullet}) + (p_{\bullet B\bullet} + p_{A\bullet\bullet} - 2p_{AB\bullet}) \\
&\qquad - 2(p_{\bullet\bullet 1} + p_{A\bullet\bullet} - 2p_{A\bullet 1}) + (2p_{A\bullet 1} + p_{\bullet\bullet\bullet} - p_{A\bullet\bullet} - p_{\bullet\bullet 1}) \\
&\qquad - 2(p_{\bullet\bullet 1} + p_{\bullet B\bullet} - 2p_{\bullet B1}) + (2p_{\bullet B1} + p_{\bullet\bullet\bullet} - p_{\bullet B\bullet} - p_{\bullet\bullet 1})
\end{aligned}$$

$$= -6(p_{AB\bullet} + p_{\bullet\bullet 1} - p_{A\bullet 1} - p_{\bullet B1}). \tag{S14}$$

Since $A \subseteq [r], B \subseteq [s]$ were arbitrary, and since $f_{\mathbb{S}} \equiv 0 \in \mathcal{G}_{\mathbb{S}}^{+}$, the desired lower bound follows.

We now give the matching upper bound on $R(P_{\mathbb{S}})$. When $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{0}$ we automatically have $R(P_{\mathbb{S}}) = 0$. On the other hand, when $P_{\mathbb{S}} \notin \mathcal{P}_{\mathbb{S}}^{0}$, we relate $R(P_{\mathbb{S}})$ to the maximum two-commodity flow through the network shown in Figure 1. Recalling the matrix $\mathbb{A} = (\mathbb{A}_{(S,y_S),x})_{(S,y_S)\in\mathcal{X}_{\mathbb{S}},x\in\mathcal{X}} \in \{0,1\}^{\mathcal{X}_{\mathbb{S}}\times\mathcal{X}}$ from (12), for any $P_{\mathbb{S}} = (P_S : S \in \mathbb{S})$ with corresponding probability mass sequence $p_{\mathbb{S}} = (p_{(S,y_S)}) \in [0,1]^{\mathcal{X}_{\mathbb{S}}}$, we may write

$$
\begin{aligned}
R(P_{\mathbb{S}}) &= -\frac{1}{|\mathbb{S}|}\min\bigl\{p_{\mathbb{S}}^T f_{\mathbb{S}} : f_{\mathbb{S}} \geq -1, \mathbb{A}^T f_{\mathbb{S}} \geq 0\bigr\} \\
&= 1 - \frac{1}{|\mathbb{S}|}\min\bigl\{p_{\mathbb{S}}^T y : y \in [0,\infty)^{\mathcal{X}_{\mathbb{S}}}, \mathbb{A}^T y \geq |\mathbb{S}| \cdot 1_{\mathcal{X}}\bigr\} \\
&= 1 - \min\bigl\{p_{\mathbb{S}}^T z : z \in [0,\infty)^{\mathcal{X}_{\mathbb{S}}}, \mathbb{A}^T z \geq 1_{\mathcal{X}}\bigr\} \\
&= 1 - \max\bigl\{1_{\mathcal{X}}^T p : p \in [0,\infty)^{\mathcal{X}}, \mathbb{A}p \leq p_{\mathbb{S}}\bigr\}. \qquad \text{(S15)}
\end{aligned}
$$

Here, the final equality follows from the strong duality theorem for linear programming (e.g. Matousek and Gärtner, 2007, p. 83), where we note that both the primal and dual problems have feasible solutions. It follows from this that

$$
\begin{aligned}
1 - R(P_{\mathbb{S}}) &= \max\bigl\{1_{\mathcal{X}}^T p : p \in [0,\infty)^{\mathcal{X}}, \mathbb{A}p \leq p_{\mathbb{S}}\bigr\}, \\
&= \max\biggl\{\sum_{i=1}^{r}\sum_{j=1}^{s}(q_{ij1} + q_{ij2}) : \min_{i,j,k} q_{ijk} \geq 0, \max_{i,j}(q_{ij1} + q_{ij2} - p_{ij\bullet}) \leq 0, \\
&\qquad \max_{i,k}\biggl(\sum_{j=1}^{s} q_{ijk} - p_{i\bullet k}\biggr) \leq 0, \max_{j,k}\biggl(\sum_{i=1}^{r} q_{ijk} - p_{\bullet jk}\biggr) \leq 0\biggr\}. \qquad \text{(S16)}
\end{aligned}
$$

Figure 1 represents a flow network where, for $k \in \{1,2\}$, commodity k is transferred from source s_k to sink t_k. We think of q_{ijk} as the flow of commodity k from node x_{ik} to node $y_{ij}^{(1)}$, and $\sum_{i=1}^{r}\sum_{j=1}^{s} q_{ijk}$ as being the total flow of commodity k from source s_k to sink t_k. Of this flow, at most $p_{i\bullet k}$ may go through x_{ik}, for each $i \in [r]$, corresponding to the constraint $\sum_{j=1}^{s} q_{ijk} \leq p_{i\bullet k}$. For each $i \in [r], j \in [s]$, the combined flow of both commodities from x_{i1} and x_{i2} through to $y_{ij}^{(2)}$ is bounded above by $p_{ij\bullet}$, corresponding to the constraint $q_{ij1} + q_{ij2} \leq p_{ij\bullet}$. For each $j \in [s]$ and $k \in \{1,2\}$, the subsequent flow of commodity k through node z_{jk} to t_k is bounded by $p_{\bullet jk}$, corresponding to the constraint $\sum_{i=1}^{r} q_{ijk} \leq p_{\bullet jk}$.

Having established the link between $R(P_{\mathbb{S}})$ and this network flow problem, we proceed to find a total flow that matches the upper bound implied by (S14) and (S16), i.e.

$$
\begin{aligned}
&1 + 2\min_{A\subseteq[r],B\subseteq[s]}(p_{AB\bullet} + p_{\bullet\bullet 1} - p_{A\bullet 1} - p_{\bullet B1}) \\
&\qquad = \min_{A\subseteq[r],B\subseteq[s]}(p_{A^c\bullet 1} + p_{A\bullet 2} + p_{\bullet B^c 1} + p_{\bullet B2} + p_{AB\bullet} + p_{A^cB^c\bullet}). \qquad \text{(S17)}
\end{aligned}
$$

The fact that the left-hand side of (S17) is equal to the right-hand side relies on the consistency of $p_{\mathbb{S}}$. Let $A \subseteq [r]$ and $B \subseteq [s]$ be minimising sets in the above display, observing that the same choices minimise both left- and right-hand sides. Then, for $i \in A$, we have

$$
p_{AB\bullet} - p_{A\bullet 1} \leq p_{A\setminus\{i\}B\bullet} - p_{A\setminus\{i\}\bullet 1} = p_{AB\bullet} - p_{iB\bullet} - p_{A\bullet 1} + p_{i\bullet 1},
$$

so that $p_{iB\bullet} \leq p_{i\bullet 1}$. It is therefore possible to send a flow of commodity 1 of $p_{ij\bullet}$ from s_1 through x_{i1} to $y_{ij}^{(2)}$, for each $(i,j) \in A \times B$. Similarly, by considering $i \in A^c$ and repeating the

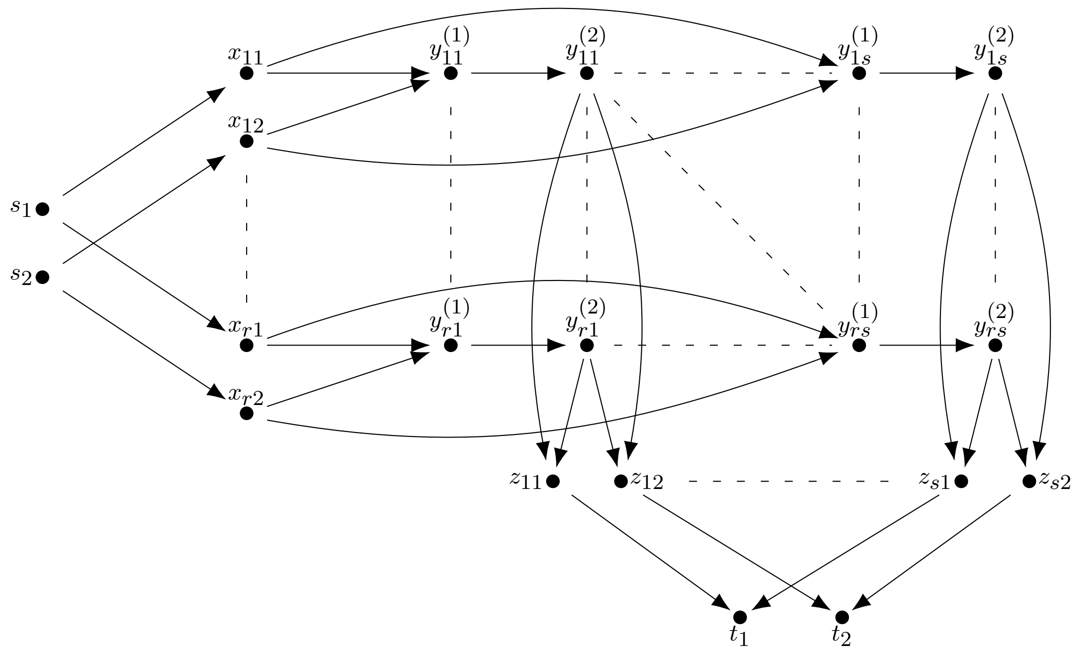

Fig 1: Illustration of the flow network described in the proof of Theorem 8. The capacity constraints are $c(s_k, x_{ik}) = p_{i\bullet k}$, $c(x_{ik}, y^{(1)}_{ij}) = \infty$, $c(y^{(1)}_{ij}, y^{(2)}_{ij}) = p_{ij\bullet}$, $c(y^{(2)}_{ij}, z_{jk}) = \infty$ and $c(z_{jk}, t_2) = p_{\bullet jk}$ for $i \in [r]$, $j \in [s]$ and $k \in [2]$.

calculation above with $A \cup \{i\}$ in place of $A \setminus \{i\}$, we see that $p_{iB^c\bullet} \le p_{i\bullet 2}$. Hence a flow of commodity 2 of $p_{ij\bullet}$ can be sent from s_2 through x_{i2} to $y^{(2)}_{ij}$ for each $(i,j) \in A^c \times B^c$. So far, then, we have shown how to send a flow of commodity 1 of $p_{AB\bullet}$ from s_1 to $\{z_{j1} : j \in B\}$, and a flow of commodity 2 of $p_{A^cB^c\bullet}$ from s_2 to $\{z_{j2} : j \in B^c\}$.

We now claim that, for each $i \in A^c$, we may send a flow of commodity 1 of $p_{i\bullet 1}$ from s_1 through x_{i1} and $y^{(2)}_{iB} := \{y^{(2)}_{ij} : j \in B\}$ to $z_{B1} := \{z_{j1} : j \in B\}$, and that this flow together with the previous flow of commodity 1 can pass through z_{B1} to t_1. To do this we use a generalisation of Hall's marriage theorem to one-commodity flows due to Gale (1957). Each z_{j1} for $j \in B$ already has an incoming flow of $p_{Aj\bullet}$, so has a remaining capacity of $p_{\bullet j1} - p_{Aj\bullet}$. By Gale's theorem, the desired flow is therefore feasible if and only if, for every $A' \subseteq A^c$ and $B' \subseteq B$, we have

$$\sum_{i \in A'} p_{i\bullet 1} - \sum_{j \in B \setminus B'} (p_{\bullet j1} - p_{Aj\bullet}) \le \sum_{i \in A'} \sum_{j \in B'} p_{ij\bullet}.$$

This condition is equivalent to the condition that, for all $A' \subseteq A^c$ and $B' \subseteq B$ we have

$$p_{(A\cup A')B'\bullet} - p_{(A\cup A')\bullet 1} - p_{\bullet B'1} \ge p_{AB\bullet} - p_{A\bullet 1} - p_{\bullet B1},$$

but we know that this is true because (A, B) are minimisers of the left-hand side of (S17). Thus, the desired flow of commodity 1 is feasible. Similarly, for each $i \in A$, we may send a flow of $p_{i\bullet 1}$ of commodity 2 from s_2 through x_{i1} and $y^{(2)}_{iB^c} := \{y^{(2)}_{ij} : j \in B^c\}$ to $z_{B^c1} := \{z_{j1} : j \in B^c\}$, and this flow can pass through to t_2. We have therefore now shown that we can send a combined flow of $p_{AB\bullet} + p_{A^cB^c\bullet} + p_{A^c\bullet 1} + p_{A\bullet 2}$ from the sources to the sinks.

Until this point, no flow has been routed through z_{B2} or z_{B^c1}. To conclude our proof, then, we now claim that it is possible to introduce an additional flow of $p_{\bullet B2}$ of commodity 2, as well as $p_{\bullet B^c 1}$ of commodity 1 into the network, to put all edges from z_{B2} to t_2 and from $z_{B^c 1}$ to t_1 at full capacity. Consider any maximal flow in the network; we wish to determine the maximal amount of commodity 2 that can be sent from s_2 through $x_{A^c 2}$ and $y^{(2)}_{A^c B}$ to z_{B2} and thus to t_2, in addition to the existing flow. To this end, suppose that there exists $j \in B$ with the edge from z_{j2} to t_2 at less than full capacity. Then, since the flow is maximal, it must be the case that for each $i \in A^c$, the flow of commodity 2 from s_1 to x_{i2} is full (i.e. equal to $p_{i\bullet 2}$), or the flow from $y^{(1)}_{ij}$ to $y^{(2)}_{ij}$ is full. However, if the flow from s_1 to x_{i2} is equal to $p_{i\bullet 2}$, then the total flow from $\{x_{i1}, x_{i2}\}$ must be equal to $p_{i\bullet 1} + p_{i\bullet 2} = p_{i\bullet\bullet} = \sum_{j'=1}^{s} p_{ij'\bullet}$. In this case, the edge from $y^{(1)}_{ij}$ to $y^{(2)}_{ij}$ must be full. So, if the edge from z_{j2} to t_2 is not full, then the edge from $y^{(1)}_{ij}$ to $y^{(2)}_{ij}$ is full for each $i \in A^c$ (and each $i \in A$ from the earlier flow). It follows that, in this case, there is a flow of $\sum_{i=1}^{r} p_{ij\bullet} = p_{\bullet j\bullet} = p_{\bullet j1} + p_{\bullet j2}$ from $y^{(1)}_{[r]j}$ to $y^{(2)}_{[r]j}$. But such a flow would put both edges z_{j1} to t_1 and z_{j2} to t_2 at full capacity, contradicting our original hypothesis. Hence, at any maximal flow, all edges from z_{B2} to t_2 are full, and similarly, all edges from $z_{B^c 1}$ to t_1 are full. Thus, we can indeed send the desired additional flow through the network, and we deduce that the total capacity of the network is at least the expression on the right-hand side of (S17). We conclude from (S16) and (S17) that

$$R(P_{\mathbb{S}}) \leq 2\max\Big\{0, \max_{A\subseteq[r], B\subseteq[s]} (-p_{AB\bullet} + p_{A\bullet 1} + p_{\bullet B1} - p_{\bullet\bullet 1})\Big\},$$

and this completes the proof of the first part of the theorem.

We now turn to the second part of our result. We first show that $p^*_{\mathbb{S}} \in \mathcal{P}^{0,*}_{\mathbb{S}} + \mathcal{P}^{\mathrm{cons},**}_{\mathbb{S}}$ if and only if $p^*_{\mathbb{S}} \in \mathcal{P}^{\mathrm{cons},*}_{\mathbb{S}}$ and

$$\max\{1^T_{\mathcal{X}} p : p \in [0,\infty)^{\mathcal{X}}, \mathbb{A}p \leq p^*_{\mathbb{S}}\} \geq (p^*_{\bullet\bullet\bullet} - 1)_+.$$

If $p^*_{\bullet\bullet\bullet} \leq 1$, then $p^*_{\mathbb{S}} \in \mathcal{P}^{\mathrm{cons},**}_{\mathbb{S}}$ and there is nothing to prove, so we assume that $p^*_{\bullet\bullet\bullet} > 1$. If $p^*_{\mathbb{S}} \in \mathcal{P}^{0,*}_{\mathbb{S}} + \mathcal{P}^{\mathrm{cons},**}_{\mathbb{S}}$, then we may write $p^*_{\mathbb{S}} = \mathbb{A}p + r_{\mathbb{S}}$ with $p \in [0,\infty)^{\mathcal{X}}$ and $r_{\mathbb{S}} \in \mathcal{P}^{\mathrm{cons},**}_{\mathbb{S}}$. Then

$$\max\{1^T_{\mathcal{X}} p' : p' \in [0,\infty)^{\mathcal{X}}, \mathbb{A}p' \leq p^*_{\mathbb{S}}\} \geq 1^T_{\mathcal{X}} p = \frac{1}{|\mathbb{S}|}\bigg(\sum_{S\in\mathbb{S}} 1_S\bigg)^T \mathbb{A}p = p^*_{\bullet\bullet\bullet} - r_{\bullet\bullet\bullet} \geq p^*_{\bullet\bullet\bullet} - 1.$$

On the other hand, suppose that $p^*_{\mathbb{S}} \in \mathcal{P}^{\mathrm{cons},*}_{\mathbb{S}}$ and that there exists $p \in [0,\infty)^{\mathcal{X}}$ with $\mathbb{A}p \leq p^*_{\mathbb{S}}$ and $1^T_{\mathcal{X}} p \geq p^*_{\bullet\bullet\bullet} - 1$. Then we certainly have $r_{\mathbb{S}} = p^*_{\mathbb{S}} - \mathbb{A}p \in \mathcal{P}^{\mathrm{cons},*}_{\mathbb{S}}$. But since we also have $r_{\bullet\bullet\bullet} = p^*_{\bullet\bullet\bullet} - 1^T_{\mathcal{X}} p \leq 1$, it follows that $r_{\mathbb{S}} \in \mathcal{P}^{\mathrm{cons},**}_{\mathbb{S}}$, and we have proved our claim. Now, the proof of the first part of the result shows that

$$\begin{aligned}\max\{1^T_{\mathcal{X}} p &: p \in [0,\infty)^{\mathcal{X}}, \mathbb{A}p \leq p^*_{\mathbb{S}}\} \\ &= \min_{A\subseteq[r], B\subseteq[s]} (p^*_{A^c\bullet 1} + p^*_{A\bullet 2} + p^*_{\bullet B^c 1} + p^*_{\bullet B2} + p^*_{AB\bullet} + p^*_{A^c B^c \bullet}) \\ &= p^*_{\bullet\bullet\bullet} + 2\min_{A\subseteq[r], B\subseteq[s]} (p^*_{AB\bullet} + p^*_{\bullet\bullet 1} - p^*_{A\bullet 1} - p^*_{\bullet B1}).\end{aligned}$$

When $p^*_{\bullet\bullet\bullet} \geq 1$, we therefore have $p^*_{\mathbb{S}} \in \mathcal{P}^{0,*}_{\mathbb{S}} + \mathcal{P}^{\mathrm{cons},**}_{\mathbb{S}}$ if and only if $p^*_{\mathbb{S}} \in \mathcal{P}^{\mathrm{cons},*}_{\mathbb{S}}$ and

$$1 + 2\min_{A\subseteq[r], B\subseteq[s]} (p^*_{AB\bullet} + p^*_{\bullet\bullet 1} - p^*_{A\bullet 1} - p^*_{\bullet B1}) \geq 0,$$

as claimed. On the other hand, when $p^*_{\bullet\bullet\bullet} < 1$ and $p^*_{\mathbb{S}} \in \mathcal{P}^{\mathrm{cons},*}_{\mathbb{S}}$, we always have $p^*_{\mathbb{S}} \in \mathcal{P}^{\mathrm{cons},**}_{\mathbb{S}} \subseteq \mathcal{P}^{0,*}_{\mathbb{S}} + \mathcal{P}^{\mathrm{cons},**}_{\mathbb{S}}$, and moreover

$$1 + 2\min_{A\subseteq[r], B\subseteq[s]} (p^*_{AB\bullet} + p^*_{\bullet\bullet 1} - p^*_{A\bullet 1} - p^*_{\bullet B1})$$

$$> p^*_{\bullet\bullet\bullet} + 2 \min_{A \subseteq [r], B \subseteq [s]} (p^*_{AB\bullet} + p^*_{\bullet\bullet 1} - p^*_{A\bullet 1} - p^*_{\bullet B1})$$
$$= \max\bigl\{1_{\mathcal{X}}^T p : p \in [0,\infty)^{\mathcal{X}}, \mathbb{A}p \le p^*_{\mathbb{S}}\bigr\} \ge 0.$$

Combining both cases, we have now shown that

$$\mathcal{P}^{0,*}_{\mathbb{S}} + \mathcal{P}^{\mathrm{cons},**}_{\mathbb{S}} = \Bigl\{ p^*_{\mathbb{S}} \in \mathcal{P}^{\mathrm{cons},*}_{\mathbb{S}} : 1 + 2 \min_{A \subseteq [r], B \subseteq [s]} (p^*_{AB\bullet} + p^*_{\bullet\bullet 1} - p^*_{A\bullet 1} - p^*_{\bullet B1}) \ge 0 \Bigr\},$$

as required. □

The proof of our lower bound in Theorem 9 relies on the following lemma, which is an extension of both Wu and Yang (2016, Lemma 3) and Jiao, Han and Weissman (2018, Lemma 32).

LEMMA S1. *Let V, V' be random variables supported on $[\lambda/2 - M, \lambda/2 + M]$ for some $M \le \lambda/2$, and suppose that $\mathbb{E}(V^\ell) = \mathbb{E}\bigl((V')^\ell\bigr)$ for $\ell \in [L]$. Let Q denote the distribution on $\mathbb{Z}^2$ of $(W_1, W_2)^T$, where, conditional on $V = v$, we have that W_1 and W_2 are independent, with $W_1 | V = v \sim \mathrm{Poi}(v)$ and $W_2 | V = v \sim \mathrm{Poi}(\lambda - v)$. Define Q' in terms of V' analogously. Then*

$$d_{\mathrm{TV}}(Q, Q') \le \frac{2^{1/2}}{\pi^{1/4}} \Bigl(\frac{2eM^2}{\lambda(L+1)} \Bigr)^{(L+1)/2}$$

whenever $L + 2 \ge 8M^2/\lambda$.

PROOF OF LEMMA S1. Let $U := (V - \lambda/2)/M$ and $U' := (V' - \lambda/2)/M$, and for $m \in \mathbb{N}$ and $x \in \mathbb{R}$, let $(x)_m := x(x-1)\dots(x-m+1)$ for the falling factorial (with $(x)_0 := 1$). Letting $Y, Z \sim \mathrm{Poi}(\lambda/2)$ be independent, we have

$$d_{\mathrm{TV}}(Q, Q') = \frac{1}{2} e^{-\lambda} \sum_{i,j=0}^{\infty} \frac{1}{i!j!} \Bigl| \mathbb{E}\bigl\{ V^i (\lambda - V)^j - (V')^i (\lambda - V')^j \bigr\} \Bigr|$$
$$= \frac{e^{-\lambda}}{2} \sum_{i,j=0}^{\infty} \frac{(\lambda/2)^{i+j}}{i!j!} \Bigl| \mathbb{E}\Bigl\{ \Bigl(1 + \frac{2MU}{\lambda}\Bigr)^i \Bigl(1 - \frac{2MU}{\lambda}\Bigr)^j - \Bigl(1 + \frac{2MU'}{\lambda}\Bigr)^i \Bigl(1 - \frac{2MU'}{\lambda}\Bigr)^j \Bigr\} \Bigr|$$
$$= \frac{1}{2} e^{-\lambda} \sum_{i,j=0}^{\infty} \frac{(\lambda/2)^{i+j}}{i!j!} \Bigl| \mathbb{E} \sum_{k=0}^{i+j} \Bigl(\frac{2M}{\lambda}\Bigr)^k \{U^k - (U')^k\} \sum_{m=0}^{k} \binom{i}{m} \binom{j}{k-m} (-1)^{k-m} \Bigr|$$
$$\le e^{-\lambda} \sum_{i,j=0}^{\infty} \frac{(\lambda/2)^{i+j}}{i!j!} \sum_{k=0}^{i+j} \Bigl(\frac{2M}{\lambda}\Bigr)^k \mathbb{1}_{\{k \ge L+1\}} \Bigl| \sum_{m=0}^{k} \binom{i}{m} \binom{j}{k-m} (-1)^{k-m} \Bigr|$$
$$= \sum_{k=L+1}^{\infty} \frac{1}{k!} \Bigl(\frac{2M}{\lambda}\Bigr)^k \mathbb{E} \Bigl| \sum_{m=0}^{k} (-1)^m \binom{k}{m} (Y)_m (Z)_{k-m} \Bigr|$$

(S18)
$$\le \sum_{k=L+1}^{\infty} \frac{1}{k!} \Bigl(\frac{2M}{\lambda}\Bigr)^k \mathbb{E}^{1/2} \Bigl[\Bigl\{ \sum_{m=0}^{k} (-1)^m \binom{k}{m} (Y)_m (Z)_{k-m} \Bigr\}^2 \Bigr].$$

We now bound this second moment using the facts that $(x)_m(x)_n = \sum_{\ell=0}^m \binom{m}{\ell}\binom{n}{\ell}\ell!(x)_{m+n-\ell}$ and $\mathbb{E}(Y)_m = (\lambda/2)^m$ for all $m,n \in \mathbb{N}_0$ to write

$$\begin{aligned}
&\mathbb{E}\biggl[\biggl\{\sum_{m=0}^k (-1)^m \binom{k}{m}(Y)_m(Z)_{k-m}\biggr\}^2\biggr] \\
&\qquad= \sum_{m,n=0}^k (-1)^{m+n}\binom{k}{m}\binom{k}{n}\mathbb{E}\{(Y)_m(Y)_n\}\mathbb{E}\{(Z)_{k-m}(Z)_{k-n}\} \\
&\qquad= \sum_{m,n=0}^k (-1)^{m+n}\binom{k}{m}\binom{k}{n}\sum_{\ell,r=0}^\infty \binom{m}{\ell}\binom{n}{\ell}\binom{k-m}{r}\binom{k-n}{r}\ell!r!(\lambda/2)^{2k-\ell-r} \\
&\qquad= \sum_{\ell,r=0}^\infty \ell!r!(\lambda/2)^{2k-\ell-r}\biggl\{\sum_{m=0}^k (-1)^m\binom{k}{m}\binom{m}{\ell}\binom{k-m}{r}\biggr\}^2. \qquad \text{(S19)}
\end{aligned}$$

Now, terms with $\ell + r > k$ are zero, because either $\ell > m$ or $r > k-m$. We can think of $\binom{m}{\ell}\binom{k-m}{r}$ as a polynomial of degree $\ell + r$ in m, and use the fact that $\sum_{m=0}^k (-1)^m\binom{k}{m}m^s = 0$ for non-negative integers $s < k$ to conclude that the only non-zero terms are those with $\ell + r = k$. We now use the fact that $\sum_{m=0}^k(-1)^m\binom{k}{m}m^k = (-1)^k k!$ to see that

$$\begin{aligned}
\sum_{m=0}^k (-1)^m\binom{k}{m}\binom{m}{\ell}\binom{k-m}{r} &= \frac{\mathbb{1}_{\{\ell+r=k\}}}{\ell!r!}\sum_{m=0}^k(-1)^m\binom{k}{m}(m)_\ell(k-m)_r \\
&= \frac{\mathbb{1}_{\{\ell+r=k\}}}{\ell!r!}(-1)^{r+k}k!. \qquad \text{(S20)}
\end{aligned}$$

From (S18), (S19) and (S20) together with Stirling's inequality (e.g. Dümbgen, Samworth and Wellner, 2021, p. 847), we deduce that when $L+2 \geq 8M^2/\lambda$, we have

$$\begin{aligned}
d_{\mathrm{TV}}(Q,Q') &\leq \sum_{k=L+1}^\infty \frac{1}{k!}\Bigl(\frac{2M}{\lambda}\Bigr)^k\biggl\{\sum_{\ell,r=0}^\infty \ell!r!(\lambda/2)^{2k-\ell-r}\frac{\mathbb{1}_{\{\ell+r=k\}}}{(\ell!)^2(r!)^2}(k!)^2\biggr\}^{1/2} \\
&= \sum_{k=L+1}^\infty \frac{1}{k!}\Bigl(\frac{2M}{\lambda}\Bigr)^k\biggl\{k!(\lambda/2)^k\sum_{\ell=0}^k\binom{k}{\ell}\biggr\}^{1/2} = \sum_{k=L+1}^\infty \frac{1}{(k!)^{1/2}}\Bigl(\frac{2M^2}{\lambda}\Bigr)^{k/2} \\
&\leq \frac{2}{\{(L+1)!\}^{1/2}}\Bigl(\frac{2M^2}{\lambda}\Bigr)^{(L+1)/2} \leq \frac{2}{\{2\pi(L+1)\}^{1/4}}\Bigl(\frac{2eM^2}{\lambda(L+1)}\Bigr)^{(L+1)/2} \\
&\leq \frac{2^{1/2}}{\pi^{1/4}}\Bigl(\frac{2eM^2}{\lambda(L+1)}\Bigr)^{(L+1)/2},
\end{aligned}$$

as required. □

PROOF OF THEOREM 9. Assume without loss of generality that $n_{\{1,2\}} \leq n_{\{1,3\}}$. We will start by showing that we may work in a Poisson sampling model without changing the separation rates. Extending our previous setting, let $(X_{S,i})_{S\in\mathbb{S}, i\in\mathbb{N}}$ denote independent random variables, with $X_{S,i} \sim P_S$, and let $N_{\mathbb{S}} := (N_S : S \in \mathbb{S})$ be an independent sequence of Poisson random variables, independent of $(X_{S,i})_{S\in\mathbb{S},i\in\mathbb{N}}$, with $\mathbb{E}N_S = n_S$ for all $S \in \mathbb{S}$. Let $\Psi'_{\mathbb{S}}$ denote the set of sequences of tests of the form $(\psi'_{n'_{\mathbb{S}}} \in \Psi_{n'_{\mathbb{S}}} : n'_{\mathbb{S}} \in \mathbb{N}_0^{\mathbb{S}})$, and write

$$\mathcal{R}^{\mathrm{Poi}}(n_{\mathbb{S}}, \rho) := \inf_{\psi'_{\mathbb{S}}\in\Psi'_{\mathbb{S}}}\biggl\{\sup_{P_{\mathbb{S}}\in\mathcal{P}^0_{\mathbb{S}}}\mathbb{E}_{P_{\mathbb{S}}}(\psi'_{N_{\mathbb{S}}}) + \sup_{P_{\mathbb{S}}\in\mathcal{P}_{\mathbb{S}}(\rho)}\mathbb{E}_{P_{\mathbb{S}}}(1-\psi'_{N_{\mathbb{S}}})\biggr\}.$$

Here, the expectations are taken over the randomness both in the data and in the sample sizes. Since $\mathcal{R}(n'_{\mathbb{S}},\rho) \le \mathcal{R}(n''_{\mathbb{S}},\delta)$ whenever $n'_S \ge n''_S$ for all $S \in \mathbb{S}$, we have that

$$\begin{aligned}
&\mathcal{R}^{\mathrm{Poi}}(n_{\mathbb{S}},\rho) \\
&\quad = \inf_{\psi'_{\mathbb{S}}\in\Psi'_{\mathbb{S}}}\biggl\{\sup_{P_{\mathbb{S}}\in\mathcal{P}^0_{\mathbb{S}}}\sum_{n'_{\mathbb{S}}\in\mathbb{N}_0^{\mathbb{S}}}\mathbb{P}(N_{\mathbb{S}}=n'_{\mathbb{S}})\mathbb{E}_{P_{\mathbb{S}}}(\psi'_{n'_{\mathbb{S}}}) + \sup_{P_{\mathbb{S}}\in\mathcal{P}_{\mathbb{S}}(\rho)}\sum_{n'_{\mathbb{S}}\in\mathbb{N}_0^{\mathbb{S}}}\mathbb{P}(N_{\mathbb{S}}=n'_{\mathbb{S}})\mathbb{E}_{P_{\mathbb{S}}}(1-\psi'_{n'_{\mathbb{S}}})\biggr\} \\
&\quad \le \inf_{\psi'_{\mathbb{S}}\in\Psi'_{\mathbb{S}}}\sum_{n'_{\mathbb{S}}\in\mathbb{N}_0^{\mathbb{S}}}\mathbb{P}(N_{\mathbb{S}}=n'_{\mathbb{S}})\biggl\{\sup_{P_{\mathbb{S}}\in\mathcal{P}^0_{\mathbb{S}}}\mathbb{E}_{P_{\mathbb{S}}}(\psi'_{n'_{\mathbb{S}}}) + \sup_{P_{\mathbb{S}}\in\mathcal{P}_{\mathbb{S}}(\rho)}\mathbb{E}_{P_{\mathbb{S}}}(1-\psi'_{n'_{\mathbb{S}}})\biggr\} \\
&\quad = \sum_{n'_{\mathbb{S}}\in\mathbb{N}_0^{\mathbb{S}}}\mathbb{P}(N_{\mathbb{S}}=n'_{\mathbb{S}})\mathcal{R}(n'_{\mathbb{S}},\rho) \\
&\quad \le \mathcal{R}(\lceil n_{\mathbb{S}}/2\rceil,\rho)\prod_{S\in\mathbb{S}}\mathbb{P}(N_S \ge \lceil n_S/2\rceil) + \sum_{S\in\mathbb{S}}\mathbb{P}(N_S < \lceil n_S/2\rceil) \\
&\quad \le \mathcal{R}(\lceil n_{\mathbb{S}}/2\rceil,\rho) + \sum_{S\in\mathbb{S}} e^{-n_S/12}.
\end{aligned}$$

Here, in the final inequality, we have used the fact that when $W \sim \mathrm{Poi}(\lambda)$, we have

$$\mathbb{P}(W-\lambda \le -x) \le e^{-\frac{x^2}{2(\lambda+x)}}$$

for all $x \ge 0$.

We will construct priors for consistent $P_{\mathbb{S}}$ over the null and alternative hypotheses that satisfy $p_{\bullet 1\bullet} = p_{\bullet\bullet 1} = 1/2$, $p_{\bullet 21} \ge 1/4$, and $p_{i\bullet\bullet} = 1/r$ and $p_{i\bullet 1} = 1/(2r)$ for each $i \in [r]$. By (8), for such $P_{\mathbb{S}}$ we have

$$\begin{aligned}
R(P_{\mathbb{S}}) &= 2\max_{j\in[2]}\biggl\{p_{\bullet j1} - \sum_{i=1}^r \min\Bigl(p_{ij\bullet},\frac{1}{2r}\Bigr)\biggr\}_+ \\
&= \max_{j\in[2]}\biggl\{2p_{\bullet j1} - \sum_{i=1}^r\biggl(p_{ij\bullet} + \frac{1}{2r} - \biggl|p_{ij\bullet} - \frac{1}{2r}\biggr|\biggr)\biggr\}_+ \\
&= \biggl\{\sum_{i=1}^r\biggl|p_{i1\bullet} - \frac{1}{2r}\biggr| - \frac{1}{2} + \max(2p_{\bullet 11} - p_{\bullet 1\bullet}, 2p_{\bullet 21} - p_{\bullet 2\bullet})\biggr\}_+ \\
&= \biggl\{\sum_{i=1}^r\biggl|p_{i1\bullet} - \frac{1}{2r}\biggr| - \frac{1}{2} + \max(1/2 - 2p_{\bullet 21}, 2p_{\bullet 21} - 1/2)\biggr\}_+ \\
&= \biggl(\sum_{i=1}^r\biggl|p_{i1\bullet} - \frac{1}{2r}\biggr| + 2p_{\bullet 21} - 1\biggr)_+.
\end{aligned}$$

We now construct our priors using results from Jiao, Han and Weissman (2018); see also Cai and Low (2011) and Wu and Yang (2016). Set $L := \lceil 2e\log r\rceil$ and let ν_0,ν_1 be probability distributions on $[-1,1]$ satisfying:

- ν_0 and ν_1 are symmetric about 0;
- $\int_{-1}^1 t^\ell\, d\nu_0(t) = \int_{-1}^1 t^\ell\, d\nu_1(t)$ for $\ell = 0,1,\ldots,L$;
- $\int_{-1}^1 |t|\, d\nu_1(t) - \int_{-1}^1 |t|\, d\nu_0(t) = 2E_L$,

where $E_L \equiv E_L\big[|\cdot|;[-1,1]\big]$ is the error in uniform norm of the best degree-L polynomial approximation to the function $x \mapsto |x|$ on $[-1,1]$. The existence of such distributions ν_0 and ν_1 follows from Jiao, Han and Weissman (2018, Lemma 29). We recall that $E_L = \beta_* L^{-1}\{1+o(1)\}$ as $L \to \infty$, where $\beta_* \approx 0.2802$ is the Bernstein constant (Bernstein, 1914). Define $g : [-1,1] \to \mathbb{R}$ by

$$g(x) := \frac{1}{r} + \delta x, \quad \text{where } \delta := \frac{1}{r} \wedge \left(\frac{\log r}{n_{\{1,2\}} r}\right)^{1/2}.$$

Further, writing $a := 1/r - \delta \geq 0$ and $b := 1/r + \delta \leq 2/r$, define distributions μ_0 and μ_1 on $[a,b]$ by $\mu_j := \nu_j \circ g^{-1}$ for $j = 0,1$. These distributions satisfy

- $\int_a^b t\, d\mu_0(t) = \int_a^b t\, d\mu_1(t) = 1/r$;
- $\int_a^b t^\ell\, d\mu_0(t) = \int_a^b t^\ell\, d\mu_1(t)$ for $\ell = 2,3,\ldots,L$;
- $\int_a^b |t - 1/r|\, d\mu_1(t) - \int_a^b |t-1/r|\, d\mu_0(t) = 2\delta E_L$.

Since $\rho^*(n_{\mathbb{S}})$ is increasing in r, we may assume without loss of generality that r is even. We will write σ_0 and σ_1 for our priors under the null and alternative hypotheses respectively. For σ_j with $j \in \{0,1\}$ and for odd $i \in [r]$, generate $2p_{i1\bullet}$ independently from μ_j. For even $i \in [r]$, set $p_{i1\bullet} := 1/r - p_{i-1,1\bullet}$ so that $p_{\bullet 1\bullet} = 1/2$ with probability one. Given $(p_{i1\bullet})_{i=1}^r$, take $p_{i2\bullet} := 1/r - p_{i1\bullet}$ and $p_{i\bullet 1} = p_{i\bullet 2} = 1/(2r)$, so that $p_{i\bullet\bullet} = 1/r$ and $p_{\bullet\bullet 1} = 1/2$. Write

$$\chi := \mathbb{E}_{\sigma_1} \sum_{i=1}^r \left| p_{i1\bullet} - \frac{1}{2r}\right| - \mathbb{E}_{\sigma_0} \sum_{i=1}^r \left| p_{i1\bullet} - \frac{1}{2r}\right| = r\delta E_L$$

and set

$$\zeta := \frac{1}{2}\mathbb{E}_{\sigma_1} \sum_{i=1}^r \left| p_{i1\bullet} - \frac{1}{2r}\right| + \frac{1}{2}\mathbb{E}_{\sigma_0} \sum_{i=1}^r \left| p_{i1\bullet} - \frac{1}{2r}\right| \leq 1/2.$$

Our prior distributions are fully specified upon choosing $p_{\bullet 21} := 1/2 - (\zeta - \chi/4)/2 \geq 1/4$. For $j \in \{0,1\}$, let

$$\Omega_{0,j} := \left\{ (-1)^j \left(\sum_{i=1}^r \left| p_{i1\bullet} - \frac{1}{2r}\right| - \mathbb{E}_{\sigma_j} \sum_{i=1}^r \left| p_{i1\bullet} - \frac{1}{2r}\right| \right) \leq \frac{\chi}{4} \right\}.$$

Then, noting that the even terms in the sum are equal to the odd terms, by Hoeffding's inequality,

$$\mathbb{P}_{\sigma_j}(\Omega_{0,j}^c) \leq \exp\left(-\frac{\chi^2}{16 r \delta^2}\right) = e^{-rE_L^2/16}.$$

Moreover, on $\Omega_{0,0}$,

$$\sum_{i=1}^r \left| p_{i1\bullet} - \frac{1}{2r}\right| + 2p_{\bullet 21} - 1$$
$$\leq \mathbb{E}_{\sigma_0}\left(\sum_{i=1}^r \left| p_{i1\bullet} - \frac{1}{2r}\right| + 2p_{\bullet 21} - 1\right) + \chi/4 = \zeta - \chi/2 - (\zeta - \chi/4) + \chi/4 = 0,$$

so that $\sigma_0\big((\mathcal{P}_{\mathbb{S}}^0)^c\big) = \mathbb{P}_{\sigma_0}\big\{R(P_{\mathbb{S}}) > 0\big\} \leq e^{-rE_L^2/16}$. On the other hand, on $\Omega_{0,1}$,

$$\sum_{i=1}^r \left| p_{i1\bullet} - \frac{1}{2r}\right| + 2p_{\bullet 21} - 1$$
$$\geq \mathbb{E}_{\mu_1}\left(\sum_{i=1}^r \left| p_{i1\bullet} - \frac{1}{2r}\right| + 2p_{\bullet 21} - 1\right) - \chi/4 = \zeta + \chi/2 - (\zeta - \chi/4) - \chi/4 = \chi/2,$$

so that $\sigma_1\big(\mathcal{P}_{\mathbb{S}}(\chi/2)^c\big) = \mathbb{P}_{\sigma_1}\big\{R(P_{\mathbb{S}}) < \chi/2\big\} \leq e^{-rE_L^2/16}$.

We finally bound the total variation distance between the marginal distributions of the data, using similar arguments to those in Wu and Yang (2016). We have

$$\begin{aligned}\mathcal{R}^{\mathrm{Poi}}(n_{\mathbb{S}}, \chi/2) &\geq \inf_{\psi'_{\mathbb{S}} \in \Psi'_{\mathbb{S}}} \Big[\mathbb{E}_{\sigma_0}\big\{\mathbb{1}_{\{P_{\mathbb{S}} \in \mathcal{P}^0_{\mathbb{S}}\}} \mathbb{E}_{P_{\mathbb{S}}}(\psi'_{N_{\mathbb{S}}})\big\} + \mathbb{E}_{\sigma_1}\big\{\mathbb{1}_{\{P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}(\chi/2)\}} \mathbb{E}_{P_{\mathbb{S}}}(1 - \psi'_{N_{\mathbb{S}}})\big\}\Big] \\ &\geq \inf_{\psi'_{\mathbb{S}} \in \Psi'_{\mathbb{S}}} \Big[\mathbb{E}_{\sigma_0}\big\{\mathbb{E}_{P_{\mathbb{S}}}(\psi'_{N_{\mathbb{S}}})\big\} + \mathbb{E}_{\sigma_1}\big\{\mathbb{E}_{P_{\mathbb{S}}}(1 - \psi'_{N_{\mathbb{S}}})\big\}\Big] - \sigma_0\big((\mathcal{P}^0_{\mathbb{S}})^c\big) - \sigma_1\big(\mathcal{P}_{\mathbb{S}}(\chi/2)^c\big) \\ &\geq 1 - d_{\mathrm{TV}}\big(\mathbb{E}_{\sigma_0} P_{\mathbb{S}}^{n_{\mathbb{S}}}, \mathbb{E}_{\sigma_1} P_{\mathbb{S}}^{n_{\mathbb{S}}}\big) - 2e^{-rE_L^2/16},\end{aligned}$$

where, for $j = 0, 1$, we write $\mathbb{E}_{\sigma_j} P_{\mathbb{S}}^{n_{\mathbb{S}}}$ for the marginal distribution of $(X_{S,i})_{S \in \mathbb{S}, i \in \mathbb{N}}$ in our Poisson model when the prior distribution for $P_{\mathbb{S}}$ is σ_j. The distributions $P_{\{2,3\}}$ and $P_{\{1,3\}}$ are deterministic and do not change between the two priors, so

$$d_{\mathrm{TV}}\big(\mathbb{E}_{\sigma_0} P_{\mathbb{S}}^{n_{\mathbb{S}}}, \mathbb{E}_{\sigma_1} P_{\mathbb{S}}^{n_{\mathbb{S}}}\big) = d_{\mathrm{TV}}\big(\mathbb{E}_{\sigma_0} P_{\{1,2\}}^{n_{\{1,2\}}}, \mathbb{E}_{\sigma_1} P_{\{1,2\}}^{n_{\{1,2\}}}\big), \tag{S21}$$

where, for $j = 0, 1$, $\mathbb{E}_{\sigma_j} P_{\{1,2\}}^{n_{\{1,2\}}}$ denotes the marginal distribution of $(X_{\{1,2\},i})_{i \in \mathbb{N}}$ in our Poisson model when the prior distribution for $(p_{i\ell\bullet})_{i \in [r], \ell \in [2]}$ is taken from the construction of σ_j. Under our Poisson sampling scheme, since $(p_{i1\bullet})_{i\,\mathrm{odd}}$ is an independent sequence, it suffices to bound the total variation distance between the distributions of random vectors (Y_1, Y_2, Y_3, Y_4) and (Z_1, Z_2, Z_3, Z_4), where $V \sim n_{\{1,2\}}\mu_0/2$, $V' \sim n_{\{1,2\}}\mu_1/2$, and with $\lambda := n_{\{1,2\}}/r$, we have

$$(Y_1, Y_2, Y_3, Y_4) | V = v \sim \mathrm{Poi}(v) \otimes \mathrm{Poi}(\lambda - v) \otimes \mathrm{Poi}(\lambda - v) \otimes \mathrm{Poi}(v)$$

for all v, and $(Z_1, Z_2, Z_3, Z_4) | V' = v \overset{\mathrm{d}}{=} (Y_1, Y_2, Y_3, Y_4) | V = v$ for all v. We now have that

$$d_{\mathrm{TV}}\big(\mathbb{E}_{\sigma_0} P_{\{1,2\}}^{n_{\{1,2\}}}, \mathbb{E}_{\sigma_1} P_{\{1,2\}}^{n_{\{1,2\}}}\big) \leq \frac{r}{2} d_{\mathrm{TV}}\big(\mathcal{L}(Y_1, Y_2, Y_3, Y_4), \mathcal{L}(Z_1, Z_2, Z_3, Z_4)\big). \tag{S22}$$

Recalling that V and V' have identical ℓth moments for $\ell \in [L]$, we have by Lemma S1 above that

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathcal{L}(Y_1, Y_2, Y_3, Y_4), \mathcal{L}(Z_1, Z_2, Z_3, Z_4)\big) \\ &\qquad = \frac{1}{2} \sum_{w,x,y,z=0}^{\infty} \frac{e^{-2\lambda}}{w!x!y!z!} \big|\mathbb{E}\big\{V^{w+z}(\lambda - V)^{x+y}\big\} - \mathbb{E}\big\{(V')^{w+z}(\lambda - V')^{x+y}\big\}\big| \\ &\qquad = \frac{1}{2} \sum_{i,j=0}^{\infty} e^{-2\lambda} \frac{1}{i!j!} \big|\mathbb{E}\big\{(2V)^i(2\lambda - 2V)^j\big\} - \mathbb{E}\big\{(2V')^i(2\lambda - 2V')^j\big\}\big| \\ &\qquad \leq \frac{2^{1/2}}{\pi^{1/4}} \Big(\frac{e \log r}{L+1}\Big)^{(L+1)/2} \end{aligned} \tag{S23}$$

since $L + 2 \geq 4 \log r$. We deduce that with $\rho = \chi/2 = r\delta E_L/2$,

$$\begin{aligned}\mathcal{R}(\lceil n_{\mathbb{S}}/2 \rceil, \rho) &\geq \mathcal{R}^{\mathrm{Poi}}(n_{\mathbb{S}}, \rho) - \sum_{S \in \mathbb{S}} e^{-n_S/12} \\ &\geq 1 - \frac{r}{2} \cdot \frac{2^{1/2}}{\pi^{1/4}} \Big(\frac{e \log r}{L+1}\Big)^{(L+1)/2} - 2e^{-rE_L^2/16} - \sum_{S \in \mathbb{S}} e^{-n_S/12} \\ &\geq 1 - \frac{r^{1 - e\log 2}}{2} \cdot \frac{2^{1/2}}{\pi^{1/4}} - 2e^{-rE_L^2/16} - \sum_{S \in \mathbb{S}} e^{-n_S/12}.\end{aligned}$$

It follows that there exists a universal constant $r_0 > 0$ such that when $\min\bigl(r, \min_{S\in\mathbb{S}} n_S\bigr) \geq r_0$ we have $\mathcal{R}(\lceil n_{\mathbb{S}}/2\rceil, \rho) \geq 1/2$, so

$$\rho^*(\lceil n_{\mathbb{S}}/2\rceil) \geq c'\biggl\{\frac{1}{\log r} \wedge \biggl(\frac{r}{(n_{\{1,2\}} \wedge n_{\{1,3\}})\log r}\biggr)^{1/2}\biggr\}$$

for some universal constant $c' > 0$. By reducing $c' > 0$ if necessary, we may therefore conclude that the same lower bound holds for $\rho^*(n_{\mathbb{S}})$.

We now prove that we always have a parametric lower bound, so that the result still holds when $2 \leq r < r_0$. Since ρ^* is increasing in r we assume without loss of generality that $r = 2$ and that $n_{\{1,2\}} = \min_{S\in\mathbb{S}} n_S$. Here we use a two-point argument. For any $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$ with $p_{1\bullet\bullet} = p_{\bullet 1\bullet} = p_{\bullet\bullet 1} = 1/2$, we have from (8) that

$$R(P_{\mathbb{S}}) = 2\max\biggl\{0, \frac{1}{2} - p_{11\bullet} - p_{\bullet 11} - p_{1\bullet 1}, p_{11\bullet} - p_{\bullet 11} + p_{1\bullet 1} - \frac{1}{2}, \\ p_{11\bullet} + p_{\bullet 11} - p_{1\bullet 1} - 1/2, -p_{11\bullet} + p_{\bullet 11} + p_{1\bullet 1} - \frac{1}{2}\biggr\}.$$

In fact, when $p_{11\bullet} + p_{\bullet 11} + p_{1\bullet 1} \leq 1/2$ we have

$$R(P_{\mathbb{S}}) = 1 - 2(p_{11\bullet} + p_{\bullet 11} + p_{1\bullet 1}).$$

Take $p_{\bullet 11} = p_{1\bullet 1} = 1/8$ so that $R(P_{\mathbb{S}}) = 1/2 - 2p_{11\bullet}$. We can therefore take $P_{\mathbb{S}}^{(0)} \in \mathcal{P}_{\mathbb{S}}^0$ to have $p_{11\bullet} = 1/4$ and $P_{\mathbb{S}}^{(1)} \in \mathcal{P}_{\mathbb{S}}\bigl((32n_{\{1,2\}})^{-1/2}\bigr)$ to have $p_{11\bullet} = 1/4 - (32n_{\{1,2\}})^{-1/2}$. We now use Pinsker's inequality to calculate that

$$\begin{aligned} d_{\mathrm{TV}}^2\bigl((P_{\mathbb{S}}^{(0)})^{n_{\mathbb{S}}}, (P_{\mathbb{S}}^{(1)})^{n_{\mathbb{S}}}\bigr) &= d_{\mathrm{TV}}^2\bigl((P_{n_{\{1,2\}}}^{(0)})^{n_{\{1,2\}}}, (P_{n_{\{1,2\}}}^{(1)})^{n_{\{1,2\}}}\bigr) \leq \frac{n_{\{1,2\}}}{2}\mathrm{KL}\bigl(P_{n_{\{1,2\}}}^{(0)}, P_{n_{\{1,2\}}}^{(1)}\bigr) \\ &= \frac{n_{\{1,2\}}}{4}\biggl\{\log\biggl(\frac{1/4}{1/4 - (32n_{\{1,2\}})^{-1/2}}\biggr) + \log\biggl(\frac{1/4}{1/4 + (32n_{\{1,2\}})^{-1/2}}\biggr)\biggr\} \\ &= \frac{n_{\{1,2\}}}{4}\log\biggl(\frac{1}{1 - 1/(2n_{\{1,2\}})}\biggr) \leq \frac{1}{4}. \end{aligned}$$

and it follows that $\rho^*(n_{\mathbb{S}}) \geq (32\min_{S\in\mathbb{S}} n_S)^{-1/2}$. By considering the different possible orderings of r, $\min_{S\in\mathbb{S}} n_S$ and r_0, we see that the claimed lower bound holds. □

PROPOSITION S2. *Let* $\mathbb{S} = \bigl\{\{1,2,3\}, \{1,3,4\}, \{1,2,4\}\bigr\}$ *with* $\mathcal{X} = [r] \times [s] \times [2] \times [2]$ *for some* $r, s \geq 2$. *There exist universal constants* $C_0, c > 0$ *such that whenever* $s \geq C_0 \log^3 r$ *we have*

$$\rho^*(n_{\mathbb{S}}) \geq c\max\biggl\{\frac{1}{\log(rs)} \wedge \biggl(\frac{rs}{(n_{\{1,2,3\}} \wedge n_{\{1,2,4\}})\log(rs)}\biggr)^{1/2}, \frac{1}{(\min_{S\in\mathbb{S}} n_S)^{1/2}}\biggr\}.$$

PROOF OF PROPOSITION S2. As in the proof of Theorem 9, we may work in a Poisson sampling model. We will construct priors σ_0 and σ_1 for $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$ under the null and alternative hypotheses respectively, that satisfy $p_{i\bullet\bullet\bullet} = 1/r$, $p_{i\bullet 1\bullet} = p_{i\bullet\bullet 1} = 1/(2r)$, $p_{i\bullet 21} \geq 1/(4r)$ for each $i \in [r]$, and $p_{ij\bullet\bullet} = 1/(rs)$ and $p_{ij\bullet 1} = 1/(2rs)$ for each $i \in [r]$ and $j \in [s]$. By Proposition 11, for such $P_{\mathbb{S}}$, we have

$$R(P_{\mathbb{S}}) = 2\sum_{i=1}^{r} \max_{A\subseteq[s]} \max_{k=1,2}(-p_{iAk\bullet} + p_{iA\bullet 1} + p_{i\bullet k1} - p_{i\bullet\bullet 1})_+$$

$$= 2\sum_{i=1}^{r}\max_{k=1,2}\biggl\{\sum_{j=1}^{s}(p_{ij\bullet 1}-p_{ijk\bullet})_{+}+p_{i\bullet k1}-p_{i\bullet\bullet 1}\biggr\}_{+}$$
$$= \sum_{i=1}^{r}\max_{k=1,2}\biggl(\sum_{j=1}^{s}|p_{ij1\bullet}-1/(2rs)|+2p_{i\bullet k1}-1/r\biggr)_{+}$$
$$= \sum_{i=1}^{r}\biggl(\sum_{j=1}^{s}|p_{ij1\bullet}-1/(2rs)|+2p_{i\bullet 21}-1/r\biggr)_{+}.$$

With $L := \lceil 2e\log(rs)\rceil$ let ν_0,ν_1 be the distributions on $[-1,1]$ defined in the proof of Theorem 9. Now, defining $g:[-1,1]\to\mathbb{R}$ by

$$g(x) := \frac{1}{s}+\delta x, \quad \text{where } \delta := \frac{1}{s}\wedge\biggl(\frac{r\log(rs)}{n_{\{1,2,3\}}s}\biggr)^{1/2},$$

set $\mu_\ell := \nu_\ell\circ g^{-1}$ for $\ell=0,1$. We will assume without loss of generality that s is even and, for each $i\in[r]$, each odd $j\in[s]$ and each $\ell=0,1$, under σ_ℓ generate $2rp_{ij1\bullet}$ independently from μ_ℓ. For $i\in[r]$ and even $j\in[s]$ set $p_{ij1\bullet}=1/(rs)-p_{i,j-1,1\bullet}$. For all $i\in[r]$ and $j\in[s]$ take $p_{ij2\bullet}=1/(rs)-p_{ij1\bullet}$ and $p_{ij\bullet 1}=p_{ij\bullet 2}=1/(2rs)$. Similarly to the proof of Theorem 9, write

$$\chi := \mathbb{E}_{\sigma_1}\sum_{i=1}^{r}\sum_{j=1}^{s}|p_{ij1\bullet}-1/(2rs)|-\mathbb{E}_{\sigma_0}\sum_{i=1}^{r}\sum_{j=1}^{s}|p_{ij1\bullet}-1/(2rs)| = s\delta E_L$$

and

$$\zeta := \frac{1}{2}\mathbb{E}_{\sigma_1}\sum_{i=1}^{r}\sum_{j=1}^{s}|p_{ij1\bullet}-1/(2rs)|+\frac{1}{2}\mathbb{E}_{\sigma_0}\sum_{i=1}^{r}\sum_{j=1}^{s}|p_{ij1\bullet}-1/(2rs)|,$$

and choose $p_{i\bullet 21}=(1/r)\{1/2-(\zeta-\chi/4)/2\}\geq 1/(4r)$ for each $i\in[r]$. Now, using a union bound and the same argument as in the proof of Theorem 9, we have

$$\mathbb{P}_{\sigma_0}(P_{\mathbb{S}}\notin\mathcal{P}^0_{\mathbb{S}})\leq r\exp\Bigl(-\frac{sE_L^2}{16}\Bigr) \quad\text{and}\quad \mathbb{P}_{\sigma_0}(R(P_{\mathbb{S}})<\chi/2)\leq r\exp\Bigl(-\frac{sE_L^2}{16}\Bigr).$$

These right-hand sides can be made arbitrarily small by choosing C_0 sufficiently large enough in our assumption that $s\geq c_0\log^3 r$. Now, as in (S21), (S22) and (S23) in the proof of Theorem 9 we use the fact that $L+2\geq 4\log(rs)$ to see that

$$d_{\mathrm{TV}}\bigl(\mathbb{E}_{\sigma_0}P_{\mathbb{S}}^{n_{\mathbb{S}}},\mathbb{E}_{\sigma_1}P_{\mathbb{S}}^{n_{\mathbb{S}}}\bigr)\leq\frac{rs}{2}\cdot\frac{2^{1/2}}{\pi^{1/4}}\biggl(\frac{e\log(rs)}{L+1}\biggr)^{(L+1)/2}\leq\frac{(rs)^{1-e\log 2}}{2^{1/2}\pi^{1/4}}.$$

The remainder of the proof is directly analogous to the proof of Theorem 9. □

PROOF OF PROPOSITION 10. Suppose that $P_{\mathbb{S}}^{-J}\in\mathcal{P}^{\mathrm{cons}}_{\mathbb{S}^{-J}}$ and let $S_1,S_2\in\mathbb{S}$ have $S_1\cap S_2\neq\emptyset$. If neither or both of S_1 and S_2 are equal to S_0, then we have immediately that $P_{S_1}^{S_1\cap S_2}=P_{S_2}^{S_1\cap S_2}$. On the other hand, if $S_1=S_0$ but $S_2\neq S_0$, say, then $P_{S_1}^{S_1\cap S_2}=P_{S_1\cap J^c}^{S_1\cap J^c\cap S_2}=P_{S_2\cap J^c}^{S_1\cap J^c\cap S_2}=P_{S_2}^{S_1\cap S_2}$. This proves the first part of the proposition.

For the second part, if $f_{\mathbb{S}-J}=(f_S:S\in\mathbb{S}^{-J})\in\mathcal{G}^+_{\mathbb{S}^{-J}}$, then we can define $f'_{\mathbb{S}}=(f'_S:S\in\mathbb{S})$ by $f'_S:=f_S$ for $S\in\mathbb{S}\setminus\{S_0\}$ and $f'_{S_0}(x_J,x_{S_0\cap J^c}):=f_{S_0\cap J^c}(x_{S_0\cap J^c})$. Then $f'_{\mathbb{S}}\in\mathcal{G}^+_{\mathbb{S}}$, and

$$R(P_{\mathbb{S}},f'_{\mathbb{S}}) = -\frac{1}{|\mathbb{S}|}\sum_{S\in\mathbb{S}}\int_{\mathcal{X}_S}f'_S(x_S)\,dP_S(x_S)$$

$$= -\frac{1}{|\mathbb{S}^{-J}|}\sum_{S\in\mathbb{S}\setminus\{S_0\}}\int_{\mathcal{X}_S} f_S(x_S)\,dP_S(x_S) - \frac{1}{|\mathbb{S}^{-J}|}\int_{\mathcal{X}_{S_0}} f_{S_0\cap J^c}(x_{S_0\cap J^c})\,dP_{S_0}(x_{S_0})$$

$$= -\frac{1}{|\mathbb{S}^{-J}|}\sum_{S\in\mathbb{S}\setminus\{S_0\}}\int_{\mathcal{X}_S} f_S(x_S)\,dP_S(x_S) - \frac{1}{|\mathbb{S}^{-J}|}\int_{\mathcal{X}_{S_0\cap J^c}} f_{S_0\cap J^c}(x_{S_0\cap J^c})\,dP_{S_0\cap J^c}(x_{S_0\cap J^c})$$

$$= R(P_{\mathbb{S}}^{-J}, f_{\mathbb{S}^{-J}}).$$

It follows that $R(P_{\mathbb{S}}) \geq R(P_{\mathbb{S}}^{-J})$. Conversely, suppose that $f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+}$ is such that $R(P_{\mathbb{S}}, f_{\mathbb{S}}) = R(P_{\mathbb{S}})$. Now define $\tilde{f}_{\mathbb{S}} = (\tilde{f}_S : S \in \mathbb{S}^{-J})$ by $\tilde{f}_S := f_S$ for $S \in \mathbb{S}\setminus\{S_0\}$ and $\tilde{f}_{S_0\cap J^c}(x_{S_0\cap J^c}) := \inf_{x'_J\in\mathcal{X}_J} f_{S_0}(x'_J, x_{S_0\cap J^c})$. Then $\tilde{f}_{\mathbb{S}} \geq -1$. Moreover, each $\tilde{f}_S$ is upper semi-continuous: this follows when $S \in \mathbb{S}\setminus\{S_0\}$ because f_S is then upper semi-continuous; on the other hand, for any $x'_J \in \mathcal{X}_J$,

$$\limsup_{x_{n,S_0\cap J^c}\to x_{S_0\cap J^c}} \tilde{f}_{S_0\cap J^c}(x_{n,S_0\cap J^c}) \leq \limsup_{x_{n,S_0\cap J^c}\to x_{S_0\cap J^c}} f_{S_0}(x'_J, x_{n,S_0\cap J^c}) \leq f_{S_0}(x'_J, x_{S_0\cap J^c}).$$

We deduce that $\limsup_{x_{n,S_0\cap J^c}\to x_{S_0\cap J^c}} \tilde{f}_{S_0\cap J^c}(x_{n,S_0\cap J^c}) \leq \tilde{f}_{S_0\cap J^c}(x_{S_0\cap J^c})$, as required. Finally, writing $\mathcal{X}_{-J} := \prod_{j\in[d]\setminus J}\mathcal{X}_j$, we have

$$\begin{aligned}\inf_{x_{-J}\in\mathcal{X}_{-J}}\sum_{S\in\mathbb{S}^{-J}}\tilde{f}_S(x_S) &= \inf_{x_{-J}\in\mathcal{X}_{-J}}\Biggl\{\sum_{S\in\mathbb{S}\setminus\{S_0\}}\tilde{f}_S(x_S) + \tilde{f}_{S_0\cap J^c}(x_{S_0\cap J^c})\Biggr\}\\ &= \inf_{x_{-J}\in\mathcal{X}_{-J}}\Biggl\{\sum_{S\in\mathbb{S}\setminus\{S_0\}} f_S(x_S) + \inf_{x_J\in\mathcal{X}_J} f_{S_0}(x_J, x_{S_0\cap J^c})\Biggr\}\\ &= \inf_{x\in\mathcal{X}}\sum_{S\in\mathbb{S}} f_S(x_S) \geq 0.\end{aligned}$$

Thus $\tilde{f}_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}^{-J}}^{+}$, and $R(P_{\mathbb{S}}^{-J}) \geq R(P_{\mathbb{S}}^{-J}, \tilde{f}_{\mathbb{S}}) \geq R(P_{\mathbb{S}}, f_{\mathbb{S}}) = R(P_{\mathbb{S}})$. □

PROOF OF PROPOSITION 11. Any $f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+}$ can be decomposed as $(f_{S|x_J} : x_J \in \mathcal{X}_J, S \in \mathbb{S})$, where $f_{S|x_J} \in \mathcal{G}_{S\cap J^c}$ is defined by $f_{S|x_J}(x_{S\cap J^c}) := f_S(x_J, x_{S\cap J^c})$. We write $f_{\mathbb{S}|x_J} := (f_{S|x_J} : S \in \mathbb{S})$. Moreover, for each $x_J \in \mathcal{X}_J$,

$$\begin{aligned}\inf_{x_{S\cap J^c}\in\mathcal{X}_{S\cap J^c}}\sum_{S\in\mathbb{S}} f_{S|x_J}(x_{S\cap J^c}) &= \inf_{x_{S\cap J^c}\in\mathcal{X}_{S\cap J^c}}\sum_{S\in\mathbb{S}} f_S(x_J, x_{S\cap J^c})\\ &\geq \inf_{x'_J\in\mathcal{X}_J, x_{S\cap J^c}\in\mathcal{X}_{S\cap J^c}}\sum_{S\in\mathbb{S}} f_S(x'_J, x_{S\cap J^c}) \geq 0,\end{aligned}$$

so $f_{\mathbb{S}|x_J} \in \mathcal{G}_{\mathbb{S}^{-J}}^{+}$ for each $x_J \in \mathcal{X}_J$. It follows that if $\epsilon > 0$, and if $f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+}$ is such that $R(P_{\mathbb{S}}, f_{\mathbb{S}}) \geq R(P_{\mathbb{S}}) - \epsilon$, then

$$\begin{aligned}R(P_{\mathbb{S}}) \leq R(P_{\mathbb{S}}, f_{\mathbb{S}}) + \epsilon &= -\frac{1}{|\mathbb{S}|}\sum_{S\in\mathbb{S}}\int_{\mathcal{X}_S} f_S(x_S)\,dP_S(x_S) + \epsilon\\ &= -\frac{1}{|\mathbb{S}|}\sum_{S\in\mathbb{S}}\int_{\mathcal{X}_J}\int_{\mathcal{X}_{S\cap J^c}} f_{S|x_J}(x_{S\cap J^c})\,dP_{S|x_J}(x_{S\cap J^c})\,dP^J(x_J) + \epsilon\\ &= \int_{\mathcal{X}_J} R(P_{\mathbb{S}|x_J}, f_{\mathbb{S}|x_J})\,dP^J(x_J) + \epsilon.\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, the desired inequality (10) follows.

Now consider the discrete case where $\mathcal{X}_j = [m_j]$ for some $m_1, \ldots, m_d \in \mathbb{N} \cup \{\infty\}$. Given any $(f_{\mathbb{S}|x_J} : x_J \in \mathcal{X}_J)$ with $f_{\mathbb{S}|x_J} \in \mathcal{G}^+_{\mathbb{S}_{-J}}$ for each $x_J \in \mathcal{X}_J$, we can define $f_{\mathbb{S}} = (f_S : S \in \mathbb{S})$ by $f_S(x_S) := f_{S|x_J}(x_{S\cap J^c})$. Then $f_S \geq -1$ for all $S \in \mathbb{S}$, each f_S is upper semi-continuous, and

$$\min_{x\in\mathcal{X}} \sum_{S\in\mathbb{S}} f_S(x_S) = \min_{x_J\in\mathcal{X}_J} \min_{x_{S\cap J^c}\in\mathcal{X}_{S\cap J^c}} \sum_{S\in\mathbb{S}} f_{S|x_J}(x_{S\cap J^c}) \geq 0.$$

Hence $f_{\mathbb{S}} \in \mathcal{G}^+_{\mathbb{S}}$. Moreover, in this discrete case, maximising $R(P_{\mathbb{S}|x_J}, \cdot)$ over $\mathcal{G}^+_{\mathbb{S}_{-J}}$ may be regarded as maximising a continuous function over a closed subset of $[-1, |\mathbb{S}| - 1]^{\mathcal{X}_{\mathbb{S}_{-J}}}$ equipped with product topology, and this is a compact set by Tychanov's theorem (e.g. Folland, 1999, Theorem 4.42). We may therefore assume that there exists $f_{\mathbb{S}|x_J} \in \mathcal{G}^+_{\mathbb{S}_{-J}}$ such that $R(P_{\mathbb{S}|x_J}, f_{\mathbb{S}|x_J}) = R(P_{\mathbb{S}|x_J})$. Then

$$\begin{aligned}
R(P_{\mathbb{S}}) \geq R(P_{\mathbb{S}}, f_{\mathbb{S}}) &= -\frac{1}{|\mathbb{S}|} \sum_{S\in\mathbb{S}} \sum_{x_J\in\mathcal{X}_J} \sum_{x_{S\cap J^c}\in\mathcal{X}_{S\cap J^c}} f_S(x_J, x_{S\cap J^c}) p_S(x_J, x_{S\cap J^c}) \\
&= \sum_{x_J\in\mathcal{X}_J} \biggl\{ -\frac{1}{|\mathbb{S}|} \sum_{S\in\mathbb{S}} \sum_{x_{S\cap J^c}\in\mathcal{X}_{S\cap J^c}} f_{S|x_J}(x_{S\cap J^c}) p_{S|x_J}(x_{S\cap J^c}) \biggr\} p^J(x_J) \\
&= \sum_{x_J\in\mathcal{X}_J} R(P_{\mathbb{S}|x_J}, f_{\mathbb{S}|x_J}) p^J(x_J) = \sum_{x_J\in\mathcal{X}_J} R(P_{\mathbb{S}|x_J}) p^J(x_J),
\end{aligned}$$

and the desired conclusion follows. □

PROOF OF PROPOSITION 12. We first establish the lower bound on $R(P_{\mathbb{S}})$. Suppose that $\epsilon \in [0,1]$ is such that $P_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}^0_{\mathbb{S}} + \epsilon\mathcal{P}_{\mathbb{S}}$. Then we can find $Q^0_{\mathbb{S}} \in \mathcal{P}^0_{\mathbb{S}}$ and $Q_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$ such that $P_{\mathbb{S}} = (1-\epsilon)Q^0_{\mathbb{S}} + \epsilon Q_{\mathbb{S}}$. But then $P_{\mathbb{S}_1} := (P_S : S \in \mathbb{S}_1)$ satisfies $P_{\mathbb{S}_1} = (1-\epsilon)Q^0_{\mathbb{S}_1} + \epsilon Q_{\mathbb{S}_1}$, so $P_{\mathbb{S}_1} \in (1-\epsilon)\mathcal{P}^0_{\mathbb{S}_1} + \epsilon\mathcal{P}_{\mathbb{S}_1}$. Hence, by Theorem 2 we have $R(P_{\mathbb{S}}) \geq R(P_{\mathbb{S}_1})$. The same argument applies to show that $R(P_{\mathbb{S}}) \geq R(P_{\mathbb{S}_2})$, and the lower bound therefore follows.

We now turn to the upper bound. For $k \in \{1,2\}$, let $I_k := \cup_{S\in\mathbb{S}_k} S$. From (S15), for each $k \in \{1,2\}$ we can find $q_k \in [0,\infty)^{\mathcal{X}_{I_k}}$ that maximises $1^T_{\mathcal{X}_{I_k}} q$ over all $q \in [0,\infty)^{\mathcal{X}_{I_k}}$ that satisfy $\mathbb{A}^k q \leq p_{\mathbb{S}_k}$, where $\mathbb{A}^k := (\mathbb{A}^k_{(S,y_S),x})_{(S,y_S)\in\mathcal{X}_{\mathbb{S}_k}, x\in\mathcal{X}_{I_k}} \in \{0,1\}^{\mathcal{X}_{\mathbb{S}_k}\times\mathcal{X}_{I_k}}$ is given by

$$\mathbb{A}^k_{(S,y_S),x} := \mathbb{1}_{\{x_S = y_S\}}.$$

Define a measure Q on $\mathcal{X}$ with mass function q given by

$$q(x) := \begin{cases} \min\bigl\{q_1^J(x_J), q_2^J(x_J)\bigr\} \cdot \frac{q_1(x_{I_1})}{q_1^J(x_J)} \cdot \frac{q_2(x_{I_2})}{q_2^J(x_J)} & \text{if } \min\bigl\{q_1^J(x_J), q_2^J(x_J)\bigr\} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then whenever $\min\{q_1^J(x_J), q_2^J(x_J)\} > 0$, we have

$$\begin{aligned}
q^J(x_J) &= \sum_{x_{J^c\cap I_1}\in\mathcal{X}_{J^c\cap I_1}} \sum_{x_{J^c\cap I_2}\in\mathcal{X}_{J^c\cap I_2}} \min\{q_1^J(x_J), q_2^J(x_J)\} \frac{q_1(x_{I_1})}{q_1^J(x_J)} \cdot \frac{q_2(x_{I_2})}{q_2^J(x_J)} \\
&= \min\{q_1^J(x_J), q_2^J(x_J)\} \sum_{x_{J^c\cap I_1}\in\mathcal{X}_{J^c\cap I_1}} \frac{q_1(x_{I_1})}{q_1^J(x_J)} \cdot \sum_{x_{J^c\cap I_2}\in\mathcal{X}_{J^c\cap I_2}} \frac{q_2(x_{I_2})}{q_2^J(x_J)} \\
&= \min\{q_1^J(x_J), q_2^J(x_J)\} = \min\bigl\{(\mathbb{A}^1 q_1)_{(J,x_J)}, (\mathbb{A}^2 q_2)_{(J,x_J)}\bigr\} \leq (p_{\mathbb{S}})_{(J,x_J)} = p_J(x_J).
\end{aligned}$$

On the other hand, if $\min\{q_1^J(x_J), q_2^J(x_J)\} = 0$, then $q^J(x_J) = 0 \leq p_S(x_S)$. Further, whenever $q_k^J(x_J) > 0$, we have for $k \in \{1,2\}$ and any $S \in \mathbb{S}_k \setminus \{J\}$ that

$$\begin{aligned} q^S(x_S) &= \sum_{x_{J\cap S^c} \in \mathcal{X}_{J\cap S^c}} \sum_{x_{J^c \cap I_k} \in \mathcal{X}_{J^c\cap I_k}} \min\{q_1^J(x_J), q_2^J(x_J)\} \frac{q_k(x_{I_k})}{q_k^J(x_J)} \\ &\leq \sum_{x_{J\cap S^c} \in \mathcal{X}_{J\cap S^c}} \sum_{x_{J^c \cap I_k} \in \mathcal{X}_{J^c\cap I_k}} q_k(x_{I_k}) \\ &= q_k^S(x_S) = (\mathbb{A}^k q_k)_{(S,x_S)} \leq (p_{\mathbb{S}})_{(S,x_S)} = p_S(x_S). \end{aligned}$$

Finally, if $q_k^J(x_J) = 0$, then $q^S(x_S) = 0 \leq p_S(x_S)$. It follows that $\mathbb{A}q \leq p_{\mathbb{S}}$, where $\mathbb{A} := (\mathbb{A}_{(S,y_S),x})_{(S,y_S)\in\mathcal{X}_{\mathbb{S}}, x\in\mathcal{X}} \in \{0,1\}^{\mathcal{X}_{\mathbb{S}}\times\mathcal{X}}$ is given by (12). Thus, from (S15),

$$\begin{aligned} R(P_{\mathbb{S}}) &\leq 1 - \sum_{x\in\mathcal{X}} q(x) = 1 - \sum_{x_J\in\mathcal{X}_J} \min\bigl\{q_1^J(x_J), q_2^J(x_J)\bigr\} \\ &= \sum_{x_J\in\mathcal{X}_J} \max\bigl\{p_J(x_J) - q_1^J(x_J), p_J(x_J) - q_2^J(x_J)\bigr\} \\ &\leq \sum_{x_J\in\mathcal{X}_J} \bigl\{p_J(x_J) - q_1^J(x_J) + p_J(x_J) - q_2^J(x_J)\bigr\} \\ &= 1 - \sum_{x_{I_1}\in\mathcal{X}_{I_1}} q_1(x_{I_1}) + 1 - \sum_{x_{I_2}\in\mathcal{X}_{I_2}} q_2(x_{I_2}) = R(P_{\mathbb{S}_1}) + R(P_{\mathbb{S}_2}), \end{aligned}$$

as required. □

PROOF OF PROPOSITION 5. Suppose that there exist $f_{\mathbb{S}} \in \mathbb{R}^{\mathcal{X}_{\mathbb{S}}}$ and $c \in \mathbb{R}$ such that $f_{\mathbb{S}}^T p_{\mathbb{S}} = c$ for all $p_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^0$. We will show that we must also have $f_{\mathbb{S}}^T p_{\mathbb{S}} = c$ for all $p_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$. In fact, by replacing $f_{\mathbb{S}}$ by $f_{\mathbb{S}} - (c/|\mathbb{S}|)1_{\mathcal{X}_{\mathbb{S}}}$, we may assume without loss of generality that $c = 0$.

In this proof we emphasise the dependence of $\mathbb{A}$ on $\mathbb{S}$ by writing $\mathbb{A}_{\mathbb{S}}$. Since $(\mathbb{A}_{\mathbb{S}}^T f_{\mathbb{S}})^T p = 0$ for all $p \in [0,1]^{\mathcal{X}}$ with $1_{\mathcal{X}}^T p = 1$, we must have that $\mathbb{A}_{\mathbb{S}}^T f_{\mathbb{S}} = 0$. We will use induction on $|\mathbb{S}|$ to deduce that $f_{\mathbb{S}}^T p_{\mathbb{S}} = 0$ for all $p_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$. When $|\mathbb{S}| = 1$, we have that if $\mathbb{A}_{\mathbb{S}}^T f_{\mathbb{S}} = 0$, then $f_{\mathbb{S}} = 0$, so $f_{\mathbb{S}}^T p_{\mathbb{S}} = 0$ for all $p_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$. As our induction hypothesis, suppose that whenever $|\mathbb{S}| \leq m$ and $f_{\mathbb{S}} \in \mathbb{R}^{\mathcal{X}_{\mathbb{S}}}$ satisfies $\mathbb{A}_{\mathbb{S}}^T f_{\mathbb{S}} = 0$, we must have $f_{\mathbb{S}}^T p_{\mathbb{S}} = 0$ for all $p_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$.

Let $\mathbb{S}$ be given with $|\mathbb{S}| = m+1$, suppose that $f_{\mathbb{S}} \in \mathbb{R}^{\mathcal{X}_{\mathbb{S}}}$ satisfies $\mathbb{A}_{\mathbb{S}}^T f_{\mathbb{S}} = 0$, and let $p_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$ be arbitrary. Without loss of generality, we may assume that $\mathcal{X}_j = [m_j]$ for $j \in [d]$ for some $m_1, \ldots, m_d \in \mathbb{N}$. Fixing $S_0 \in \mathbb{S}$, we have

$$f_{S_0}(x_{S_0}) = -\sum_{S\in\mathbb{S}\setminus\{S_0\}} f_S(x_{S_0\cap S}, 1_{S_0^c\cap S})$$

for all $x_{S_0} \in \mathcal{X}_{S_0}$, since $(\mathbb{A}_{\mathbb{S}}^T f_{\mathbb{S}})_{(x_{S_0}, 1_{[d]\setminus S_0})} = 0$. Using the notational convention that $\sum_{x_{S_1\cap S_2}\in\mathcal{X}_{S_1\cap S_2}} p_{S_1}^{S_1\cap S_2}(x_{S_1\cap S_2}) = 1$ whenever $S_1 \cap S_2 = \emptyset$, we may therefore write

$$\begin{aligned} f_{\mathbb{S}}^T p_{\mathbb{S}} &= \sum_{x_{S_0}\in\mathcal{X}_{S_0}} f_{S_0}(x_{S_0}) p_{S_0}(x_{S_0}) + \sum_{S\in\mathbb{S}\setminus\{S_0\}} \sum_{x_S\in\mathcal{X}_S} f_S(x_S) p_S(x_S) \\ &= \sum_{S\in\mathbb{S}\setminus\{S_0\}} \biggl\{ \sum_{x_S\in\mathcal{X}_S} f_S(x_S) p_S(x_S) - \sum_{x_{S_0\cap S}\in\mathcal{X}_{S_0\cap S}} f_S(x_{S_0\cap S}, 1_{S_0^c\cap S}) p_{S_0}^{S_0\cap S}(x_{S_0\cap S}) \biggr\} \end{aligned}$$

$$= \sum_{S\in\mathbb{S}\setminus\{S_0\}} \biggl\{ \sum_{x_S\in\mathcal{X}_S} f_S(x_S)p_S(x_S) - \sum_{x_{S_0\cap S}\in\mathcal{X}_{S_0\cap S}} f_S(x_{S_0\cap S}, 1_{S_0^c\cap S})p_S^{S_0\cap S}(x_{S_0\cap S}) \biggr\}$$

$$= \sum_{S\in\mathbb{S}\setminus\{S_0\}} \sum_{x_S\in\mathcal{X}_S} p_S(x_S)\bigl\{f_S(x_S) - f_S(x_{S_0\cap S}, 1_{S_0^c\cap S})\bigr\} = (\bar{f}_{\mathbb{S}\setminus\{S_0\}})^T p_{\mathbb{S}\setminus\{S_0\}}, \tag{S24}$$

where we define $\bar{f}_{\mathbb{S}\setminus\{S_0\}} \in \mathbb{R}^{\mathcal{X}_{\mathbb{S}\setminus\{S_0\}}}$ by $\bar{f}_S(x_S) := f_S(x_S) - f_S(x_{S_0\cap S}, 1_{S_0^c\cap S})$ for $S \in \mathbb{S} \setminus \{S_0\}$ and $x_S \in \mathcal{X}_S$, and where $p_{\mathbb{S}\setminus\{S_0\}} := (p_S : S \in \mathbb{S} \setminus \{S_0\})$. For any $x \in \mathcal{X}$, we have

$$(\mathbb{A}_{\mathbb{S}\setminus\{S_0\}}^T \bar{f}_{\mathbb{S}\setminus\{S_0\}})_x = \sum_{S\in\mathbb{S}\setminus\{S_0\}} \bar{f}_S(x_S) = \sum_{S\in\mathbb{S}\setminus\{S_0\}} f_S(x_S) - \sum_{S\in\mathbb{S}\setminus\{S_0\}} f_S(x_{S_0\cap S}, 1_{S_0^c\cap S})$$

$$= (\mathbb{A}_{\mathbb{S}}^T f_{\mathbb{S}})_x - f_{S_0}(x_{S_0}) - \bigl\{(\mathbb{A}_{\mathbb{S}}^T f_{\mathbb{S}})_{(x_{S_0}, 1_{[d]\setminus\{S_0\}})} - f_{S_0}(x_{S_0})\bigr\} = f_{S_0}(x_{S_0}) - f_{S_0}(x_{S_0}) = 0.$$

Since $p_{\mathbb{S}\setminus\{S_0\}}$ satisfies the consistency constraints associated with $\mathbb{S}$, we see by (S24) and our induction hypothesis that

$$f_{\mathbb{S}}^T p_{\mathbb{S}} = (\bar{f}_{\mathbb{S}\setminus\{S_0\}})^T p_{\mathbb{S}\setminus\{S_0\}} = 0,$$

as required. □

PROPOSITION S3. *Suppose that* $\mathbb{S} = \bigl\{\{1,2\},\{2,3\},\{3,4\},\{1,4\}\bigr\}$, $\mathcal{X}_1 = [r]$ *for some* $r \in \mathbb{N}$, *and* $\mathcal{X}_2 = \mathcal{X}_3 = \mathcal{X}_4 = [2]$. *Then*

$$R(P_{\mathbb{S}}) = 2 \max_{k,\ell\in[2]} \biggl\{ p_{\bullet\bullet k\ell} - p_{\bullet 2k\bullet} - \sum_{i=1}^r \min(p_{i1\bullet\bullet}, p_{i\bullet\bullet\ell}) \biggr\}_+ . \tag{S25}$$

PROOF OF PROPOSITION S3. We first prove that $R(P_{\mathbb{S}})$ is bounded below by the quantity on the right-hand side of (S25), before proving the corresponding upper bound. First, we always have $R(P_{\mathbb{S}}) \geq 0$. Now, define $f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}$ by setting, for $i \in [r]$,

$$(f_{i1\bullet\bullet}, f_{i2\bullet\bullet}, f_{i\bullet\bullet1}, f_{i\bullet\bullet2}) := \begin{cases} (3,-1,-1,3) \text{ if } p_{i1\bullet\bullet} \leq p_{i\bullet\bullet1} \\ (-1,3,3,-1) \text{ otherwise} \end{cases},$$

$f_{\bullet\bullet12} = f_{\bullet\bullet21} = f_{\bullet12\bullet} = f_{\bullet21\bullet} = 3$ and $f_{\bullet\bullet11} = f_{\bullet\bullet22} = f_{\bullet11\bullet} = f_{\bullet22\bullet} = -1$. It is straightforward to check that $f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+$. Now

$$R(P_{\mathbb{S}}, f_{\mathbb{S}})$$

$$= -\frac{1}{4}\sum_{i=1}^r \biggl(\sum_{j=1}^2 p_{ij\bullet\bullet} f_{ij\bullet\bullet} + \sum_{\ell=1}^2 p_{i\bullet\bullet\ell} f_{i\bullet\bullet\ell} \biggr) - \frac{1}{4}\sum_{j,k=1}^2 p_{\bullet jk\bullet} f_{\bullet jk\bullet} - \frac{1}{4}\sum_{k,\ell=1}^2 p_{\bullet\bullet k\ell} f_{\bullet\bullet k\ell}$$

$$= -\frac{1}{4}\sum_{i=1}^r \{3\min(p_{i1\bullet\bullet}, p_{i\bullet\bullet1}) - \max(p_{i2\bullet\bullet}, p_{i\bullet\bullet2}) - \max(p_{i1\bullet\bullet}, p_{i\bullet\bullet1}) + 3\min(p_{i2\bullet\bullet}, p_{i\bullet\bullet2})\}$$
$$\quad - \frac{1}{4}\{3(p_{\bullet12\bullet} + p_{\bullet21\bullet}) - (p_{\bullet11\bullet} + p_{\bullet22\bullet})\} - \frac{1}{4}\{3(p_{\bullet\bullet12} + p_{\bullet\bullet21}) - (p_{\bullet\bullet11} + p_{\bullet\bullet22})\}$$

$$= -\frac{1}{4}\sum_{i=1}^r \{4\min(p_{i1\bullet\bullet}, p_{i\bullet\bullet1}) - 4\max(p_{i1\bullet\bullet}, p_{i\bullet\bullet1}) + 2p_{i\bullet\bullet\bullet}\}$$
$$\quad - \frac{1}{4}\{3(2p_{\bullet21\bullet} - p_{\bullet2\bullet\bullet} - p_{\bullet\bullet1\bullet} + p_{\bullet\bullet\bullet\bullet}) - (p_{\bullet\bullet1\bullet} - 2p_{\bullet21\bullet} + p_{\bullet2\bullet\bullet})\}$$
$$\quad - \frac{1}{4}\{3(p_{\bullet\bullet1\bullet} - 2p_{\bullet\bullet11} + p_{\bullet\bullet\bullet1}) - (2p_{\bullet\bullet11} - p_{\bullet\bullet1\bullet} - p_{\bullet\bullet\bullet1} + p_{\bullet\bullet\bullet\bullet})\}$$

$$
\begin{aligned}
&= -\frac{1}{4}\sum_{i=1}^{r}\{8\min(p_{i1\bullet\bullet}, p_{i\bullet\bullet1}) - 4p_{i1\bullet\bullet} - 4p_{i\bullet\bullet1} + 2p_{i\bullet\bullet\bullet}\} \\
&\qquad - \frac{1}{4}\big(8p_{\bullet21\bullet} - 4p_{\bullet2\bullet\bullet} - 4p_{\bullet\bullet1\bullet} + 3p_{\bullet\bullet\bullet\bullet}\big) - \frac{1}{4}\big(-8p_{\bullet\bullet11} + 4p_{\bullet\bullet1\bullet} + 4p_{\bullet\bullet\bullet1} - p_{\bullet\bullet\bullet\bullet}\big) \\
&= 2\bigg\{p_{\bullet\bullet11} - p_{\bullet21\bullet} - \sum_{i=1}^{r}\min(p_{i1\bullet\bullet}, p_{i\bullet\bullet1})\bigg\}.
\end{aligned}
$$

Since $R(P_{\mathbb{S}}) \geq R(P_{\mathbb{S}}, f_{\mathbb{S}})$, this completes the lower bound in the case that $(k,\ell) = (1,1)$ is the maximiser in (S25). The other three cases follow by almost identical arguments by choosing different $f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+}$ appropriately. We now turn to the upper bound, which we will prove by using the dual formulation

$$
1 - R(P_{\mathbb{S}}) = \max\bigg\{\sum_{i=1}^{r}\sum_{j,k,\ell=1}^{2} q_{ijk\ell} : q \in [0,\infty)^{\mathcal{X}}, \mathbb{A}q \leq p_{\mathbb{S}}\bigg\}.
$$

Write $A := \{i \in [r] : p_{i1\bullet\bullet} \leq p_{i\bullet\bullet1}\}$ and suppose that

(S26) $$p_{\bullet\bullet11} - p_{\bullet21\bullet} - p_{A1\bullet\bullet} - p_{A^c\bullet\bullet1} \geq 0,$$

where we note that an alternative expression for the left-hand side of (S26) is given by $p_{\bullet\bullet11} - p_{\bullet21\bullet} - \sum_{i=1}^{r}\min(p_{i1\bullet\bullet}, p_{i\bullet\bullet1})$. For $i \in [r]$, consider the choices

$$
\begin{aligned}
q_{i111} &= \min(p_{i1\bullet\bullet}, p_{i\bullet\bullet1}), & q_{i112} &= \frac{(p_{i1\bullet\bullet} - p_{i\bullet\bullet1})_+}{p_{A^c1\bullet\bullet} - p_{A^c\bullet\bullet1}}p_{\bullet\bullet12}, \\
q_{i121} &= 0, & q_{i122} &= \frac{(p_{i1\bullet\bullet} - p_{i\bullet\bullet1})_+}{p_{A^c1\bullet\bullet} - p_{A^c\bullet\bullet1}}p_{\bullet12\bullet}, \\
q_{i222} &= \min(p_{i\bullet\bullet2}, p_{i2\bullet\bullet}), & q_{i211} &= \frac{(p_{i\bullet\bullet1} - p_{i1\bullet\bullet})_+}{p_{A\bullet\bullet1} - p_{A1\bullet\bullet}}p_{\bullet21\bullet}, \\
q_{i212} &= 0, & q_{i221} &= \frac{(p_{i\bullet\bullet1} - p_{i1\bullet\bullet})_+}{p_{A\bullet\bullet1} - p_{A1\bullet\bullet}}p_{\bullet\bullet21},
\end{aligned}
$$

where we interpret $q_{i211} = q_{1221} = 0$ if $p_{A\bullet\bullet1} = p_{A1\bullet\bullet}$. It is clear that $q \in [0,\infty)^{\mathcal{X}}$, and we now check that $\mathbb{A}q \leq p_{\mathbb{S}}$. First,

$$
\begin{aligned}
\sum_{k,\ell=1}^{2} q_{i1k\ell} &= \min(p_{i1\bullet\bullet}, p_{i\bullet\bullet1}) + \frac{(p_{i1\bullet\bullet} - p_{i\bullet\bullet1})_+}{p_{A^c1\bullet\bullet} - p_{A^c\bullet\bullet1}}(p_{\bullet\bullet12} + p_{\bullet12\bullet}) \\
&= \min(p_{i1\bullet\bullet}, p_{i\bullet\bullet1}) + \frac{(p_{i1\bullet\bullet} - p_{i\bullet\bullet1})_+}{p_{A^c1\bullet\bullet} - p_{A^c\bullet\bullet1}}(p_{\bullet21\bullet} - p_{\bullet\bullet11} + p_{\bullet1\bullet\bullet}) \\
&\leq \min(p_{i1\bullet\bullet}, p_{i\bullet\bullet1}) + (p_{i1\bullet\bullet} - p_{i\bullet\bullet1})_+ = p_{i1\bullet\bullet},
\end{aligned}
$$

for each $i \in [r]$, where the inequality follows from (S26). It is very similar to check that $\sum_{k,\ell=1}^{2} q_{i2k\ell} \leq p_{i2\bullet\bullet}$, that $\sum_{j,k=1}^{2} q_{ijk1} \leq p_{i\bullet\bullet1}$, and that $\sum_{j,k=1}^{2} q_{ijk2} \leq p_{i\bullet\bullet2}$ for each $i \in [r]$. Now

$$
\begin{aligned}
\sum_{i=1}^{r}\sum_{\ell=1}^{2} q_{i11\ell} &= \sum_{i=1}^{r}\bigg\{\min(p_{i1\bullet\bullet}, p_{i\bullet\bullet1}) + \frac{(p_{i1\bullet\bullet} - p_{i\bullet\bullet1})_+}{p_{A^c1\bullet\bullet} - p_{A^c\bullet\bullet1}}p_{\bullet\bullet12}\bigg\} \\
&= p_{A1\bullet\bullet} + p_{A^c\bullet\bullet1} + p_{\bullet\bullet12} \leq p_{\bullet\bullet11} - p_{\bullet21\bullet} + p_{\bullet\bullet12} = p_{\bullet11\bullet},
\end{aligned}
$$

where the inequality again follows from (S26). It is similar to check that $\sum_{i=1}^r \sum_{\ell=1}^2 q_{i22\ell} \leq p_{\bullet 22\bullet}$, that $\sum_{i=1}^r \sum_{j=1}^2 q_{ij11} \leq p_{\bullet\bullet 11}$, and that $\sum_{i=1}^r \sum_{j=1}^2 q_{ij22} \leq p_{\bullet\bullet 22}$. Finally, using similar arguments we see that $\sum_{i=1}^r \sum_{\ell=1}^2 q_{i12\ell} = p_{\bullet 12\bullet}$, that $\sum_{i=1}^r \sum_{\ell=1}^2 q_{i21\ell} = p_{\bullet 21\bullet}$, that $\sum_{i=1}^r \sum_{j=1}^2 q_{ij21} = p_{\bullet\bullet 21}$, and that $\sum_{i=1}^r \sum_{j=1}^2 q_{ij12} = p_{\bullet\bullet 12}$. Now that we have seen that q satisfies the necessary constraints, we calculate that

$$\begin{aligned} R(P_{\mathbb{S}}) &\leq 1 - \sum_{i=1}^r \sum_{j,k,\ell=1}^2 q_{ijk\ell} \\ &= 1 - (p_{A1\bullet\bullet} + p_{A^c\bullet\bullet 1} + p_{\bullet\bullet 12} + p_{\bullet 12\bullet} + p_{\bullet 21\bullet} + p_{\bullet\bullet 21} + p_{A^c 2\bullet\bullet} + p_{A\bullet\bullet 2}) \\ &= 2\biggl\{ p_{\bullet\bullet 11} - p_{\bullet 21\bullet} - \sum_{i=1}^r \min(p_{i1\bullet\bullet}, p_{i\bullet\bullet 1}) \biggr\}. \end{aligned}$$

This deals with the case where $(k,\ell) = (1,1)$ gives the maximiser in (S25) and where the right-hand side of (S25) is positive, as in this case (S26) must hold. The other cases follow by very similar arguments, and this completes the proof. □

PROOF OF THEOREM 15. Given $S \in \mathbb{S}$ and $k = (k_1, \ldots, k_d) \in \mathcal{K}_h$, we can define a discretised version Q_S of P_S with mass function

$$q_S(k_S) := P_S\biggl(\prod_{j \in S \in [d_0]} I_{h_j,k_j} \times \prod_{j \in S \cap ([d]\setminus[d_0])} \{k_j\} \biggr).$$

Then $R_h(\widehat{P}_{\mathbb{S}}) \overset{\mathrm{d}}{=} R(\widehat{Q}_{\mathbb{S}})$, where $(Y_{S,i} : S \in \mathbb{S}, i \in [n_S])$ are independent with $Y_{S,i} \sim Q_S$ for $i \in [n_S]$, and $\widehat{Q}_{\mathbb{S}}$ denotes their empirical distribution. Moreover, if $R(P_{\mathbb{S}}) = 0$, then $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^0$ so there exists a distribution P on $\mathcal{X}$ whose marginal distribution on $\mathcal{X}_S$ is P_S, for each $S \in \mathbb{S}$. The discretised version Q of P with mass function

$$q(k) := P\biggl(\prod_{j=1}^{d_0} I_{h_j,k_j} \times \prod_{j=d_0+1}^{d} \{k_j\} \biggr) \tag{S27}$$

on $\mathcal{K}_h$ then satisfies the condition that its marginal on $(\mathcal{K}_h)_S$ is q_S, for each $S \in \mathbb{S}$. It follows that Q is compatible, i.e. $R(Q_{\mathbb{S}}) = 0$. The Type I error probability control follows from this and the first parts of Theorems 4 and 7.

For the second claim, given $\epsilon > 0$, find $f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+$ with $R(P_{\mathbb{S}}, f_{\mathbb{S}}) \geq R(P_{\mathbb{S}}) - \epsilon$. As in the proof of Proposition 6, we may assume without loss of generality that $f_S \leq |\mathbb{S}| - 1$ for all $S \in \mathbb{S}$. Now define $f_{\mathbb{S},h} = (f_{S,h} : S \in \mathbb{S})$ by

$$f_{S,h}(x_{S\cap[d_0]}, x_{S\cap([d]\setminus[d_0])}) := \frac{\int_{\prod_{j\in S\cap[d_0]} I_{h_j,k_j}} f_S(x'_{S\cap[d_0]}, x_{S\cap([d]\setminus[d_0])})\, dx'_{S\cap[d_0]}}{\int_{\prod_{j\in S\cap[d_0]} I_{h_j,k_j}} dx'_{S\cap[d_0]}}$$

for $(x_{S\cap[d_0]}, x_{S\cap([d]\setminus[d_0])}) \in \mathcal{X}_S$ with $x_{S\cap[d_0]} \in \prod_{j\in S\cap[d_0]} I_{h_j,k_j}$. Each $f_{S,h}$ is then clearly piecewise constants on the appropriate sets, and is bounded below by -1. To check the other constraints of $\mathcal{G}_{\mathbb{S},h}^+$, let $x \in \mathcal{X}$ be given and let U be uniformly distributed on the part of the partition of $[0,1)^{S\cap[d_0]}$ that contains $x_{[d_0]}$. We have that

$$\sum_{S\in\mathbb{S}} f_{S,h}(x_S) = \sum_{S\in\mathbb{S}} \mathbb{E}\bigl\{ f_S(U_{S\cap[d_0]}, x_{S\cap([d]\setminus[d_0])}) \bigr\} = \mathbb{E}\biggl\{ \sum_{S\in\mathbb{S}} f_S(U_{S\cap[d_0]}, x_{S\cap([d]\setminus[d_0])}) \biggr\} \geq 0,$$

and thus indeed $f_{\mathbb{S},h} \in \mathcal{G}^+_{\mathbb{S},h}$. Now,

$$R(P_{\mathbb{S}}, f_{\mathbb{S},h}) = -\frac{1}{|\mathbb{S}|}\sum_{S\in\mathbb{S}}\sum_{k\in(\mathcal{K}_h)_S} \frac{\int_{\prod_{j\in S\cap[d_0]} I_{h_j,k_j}} f_S(x'_{S\cap[d_0]}, k_{S\cap([d]\setminus[d_0])})\,dx'_{S\cap[d_0]}}{\int_{\prod_{j\in S\cap[d_0]} I_{h_j,k_j}} dx'_{S\cap[d_0]}} \times P_S\biggl(\prod_{j\in S\cap[d_0]} I_{h_j,k_j} \times \prod_{j\in S\cap([d]\setminus[d_0])}\{k_j\}\biggr)$$

$$\geq -\frac{1}{|\mathbb{S}|}\sum_{S\in\mathbb{S}}\sum_{k\in(\mathcal{K}_h)_S}\int_{\prod_{j\in S\cap[d_0]} I_{h_j,k_j}} f_S(x'_{S\cap[d_0]}, k_{S\cap([d]\setminus[d_0])})\,dP_S(x'_{S\cap[d_0]}, k_{S\cap([d]\setminus[d_0])})$$

$$- \frac{L(|\mathbb{S}|-1)}{|\mathbb{S}|}\biggl(\sum_{j=1}^{d_0} h_j^{r_j}\biggr)\sum_{S\in\mathbb{S}}\sum_{k\in(\mathcal{K}_h)_S} p_S^{S\cap([d]\setminus[d_0])}(k_{S\cap([d]\setminus[d_0])})\int_{\prod_{j\in S\cap[d_0]} I_{h_j,k_j}} dx'_{S\cap[d_0]}$$

$$= R(P_{\mathbb{S}}, f_{\mathbb{S}}) - L(|\mathbb{S}|-1)\sum_{j=1}^{d_0} h_j^{r_j} \geq R(P_{\mathbb{S}}) - \epsilon - L(|\mathbb{S}|-1)\sum_{j=1}^{d_0} h_j^{r_j}.$$

Since $\epsilon > 0$ was arbitrary, we deduce that

$$R_h(P_{\mathbb{S}}) \geq R(P_{\mathbb{S}}) - L(|\mathbb{S}|-1)\sum_{j=1}^{d_0} h_j^{r_j}. \tag{S28}$$

The completion of the argument is now very similar to the first part of the theorem: we define the discretised version Q_S of P_S via (S27). Note again that $R_h(\widehat{P}_{\mathbb{S}}) \overset{\mathrm{d}}{=} R(\widehat{Q}_{\mathbb{S}})$, where $(Y_{S,i} : S \in \mathbb{S}, i \in [n_S])$ are independent with $Y_{S,i} \sim Q_S$ for $i \in [n_S]$, and $\widehat{Q}_{\mathbb{S}}$ denotes their empirical distribution. Since $R(Q_{\mathbb{S}}) = R_h(P_{\mathbb{S}})$, the result follows from (S28) together with the second parts of Theorems 4 and 7. □

PROOF OF PROPOSITION 16. To prove the first claim, let $P_{\mathbb{S}} \in (\mathcal{P}^0_{\mathbb{S}})^{-C_\alpha}$, so that

$$\mathbb{P}_{P_{\mathbb{S}}}\biggl(1 + \sum_{b=1}^{B} \mathbb{1}_{\{R(\widehat{Q}^{(b)}_{\mathbb{S}}) \geq R(\widehat{P}_{\mathbb{S}})\}} \leq \alpha(B+1)\biggr) \leq \mathbb{P}_{P_{\mathbb{S}}}(R(\widehat{P}_{\mathbb{S}}) > 0) = \mathbb{P}_{P_{\mathbb{S}}}(\widehat{P}_{\mathbb{S}} \notin \mathcal{P}^0_{\mathbb{S}})$$
$$\leq \mathbb{P}_{P_{\mathbb{S}}}\biggl(\sum_{S\in\mathbb{S}} d_{\mathrm{TV}}(\widehat{P}_S, P_S) > C_\alpha\biggr) \leq \alpha,$$

where the final inequality follows by very similar arguments to those used to prove Proposition 4.

For the second bound, for any $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$ we may use Markov's inequality and our lower bound on B to see that

$$\mathbb{P}_{P_{\mathbb{S}}}\biggl(1 + \sum_{b=1}^{B} \mathbb{1}_{\{R(\widehat{Q}^{(b)}_{\mathbb{S}}) \geq R(\widehat{P}_{\mathbb{S}})\}} > \alpha(B+1)\biggr) \leq \frac{B\mathbb{P}_{P_{\mathbb{S}}}\bigl(R(\widehat{Q}^{(1)}_{\mathbb{S}}) \geq R(\widehat{P}_{\mathbb{S}})\bigr)}{\alpha(B+1)-1}$$
$$\leq \frac{2}{\alpha}\mathbb{P}_{P_{\mathbb{S}}}\bigl(R(\widehat{Q}^{(1)}_{\mathbb{S}}) \geq R(\widehat{P}_{\mathbb{S}})\bigr).$$

Now, if $R(P_{\mathbb{S}}) \geq \epsilon = 2C_\delta$ for some $\delta \in (0,1)$, then

$$\mathbb{P}_{P_{\mathbb{S}}}\bigl(R(\widehat{Q}^{(1)}_{\mathbb{S}}) \geq R(\widehat{P}_{\mathbb{S}})\bigr) \leq \mathbb{P}_{P_{\mathbb{S}}}\bigl(R(\widehat{Q}^{(1)}_{\mathbb{S}}) \geq \epsilon/2\bigr) + \mathbb{P}_{P_{\mathbb{S}}}\bigl(R(\widehat{P}_{\mathbb{S}}) < \epsilon/2\bigr)$$
$$\leq \mathbb{P}_{P_{\mathbb{S}}}\bigl(R(\widehat{Q}^{(1)}_{\mathbb{S}}) - R(\widehat{Q}_{\mathbb{S}}) \geq \epsilon/2\bigr) + \mathbb{P}_{P_{\mathbb{S}}}\bigl(R(\widehat{P}_{\mathbb{S}}) - R(P_{\mathbb{S}}) < -\epsilon/2\bigr)$$

$$\leq \sup_{P'_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}} \mathbb{P}_{P'_{\mathbb{S}}}\big(R(\widehat{P}'_{\mathbb{S}}) - R(P'_{\mathbb{S}}) \geq \epsilon/2\big) + \sup_{P'_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}} \mathbb{P}_{P'_{\mathbb{S}}}\big(R(\widehat{P}'_{\mathbb{S}}) - R(P'_{\mathbb{S}}) \leq -\epsilon/2\big) \leq 2\delta,$$

where $\widehat{P}'_{\mathbb{S}}$ denotes the family of empirical distributions of independent samples of sizes $n_{\mathbb{S}}$ from $P'_{\mathbb{S}}$, and where the final inequality again follows from almost identical arguments to those used in the proof of Proposition 4. We choose $\delta = \alpha\beta/4$ and complete the proof on noting that

$$2C_{\alpha\beta/4} = \sum_{S \in \mathbb{S}} \Big(\frac{|\mathcal{X}_S| - 1}{n_S}\Big)^{1/2} + \biggl\{2\log\Big(\frac{4}{\alpha\beta}\Big) \sum_{S \in \mathbb{S}} \frac{1}{n_S}\biggr\}^{1/2} \leq 2\sqrt{2}(C_\alpha + C_\beta),$$

as required. □

S2. Glossary of topological definitions. A topological space $\mathcal{X}$ is said to be *completely regular* if for every closed set $B \subseteq \mathcal{X}$ and and every $x_0 \in \mathcal{X} \setminus B$, there exists a bounded continuous function $f : \mathcal{X} \to \mathbb{R}$ such that $f(x_0) = 1$ and $f(x) = 0$ for all $x \in B$. We say $\mathcal{X}$ is *Hausdorff* if, given any distinct $x, y \in \mathcal{X}$, there exist open sets $U \subseteq \mathcal{X}$ containing x and $V \subseteq \mathcal{X}$ such that $U \cap V = \emptyset$. We say a subset of $\mathcal{X}$ is *σ-compact* if it is countable union of compact sets. Given a Borel subset E of $\mathcal{X}$, we say a Borel measure μ on $\mathcal{X}$ is *outer regular* on E if

$$\mu(E) = \inf\{\mu(U) : U \supseteq E, U \text{ open}\}$$

and *inner regular* on E if

$$\mu(E) = \sup\{\mu(K) : K \subseteq E, K \text{ compact}\}.$$

We say μ is a *Radon* measure if it is outer regular on all Borel sets, inner regular on all open sets, and finite on all compact sets.

If $\mathcal{T}$ is a topology on $\mathcal{X}$, a *neighbourhood base* for $\mathcal{T}$ at $x \in \mathcal{X}$ is a family $\mathcal{N} \subseteq \mathcal{T}$ such that $x \in V$ for all $V \in \mathcal{N}$ and, whenever $U \in \mathcal{T}$ and $x \in U$, there exists $V \in \mathcal{N}$ such that $V \subseteq U$. A *base* for $\mathcal{T}$ is a family $\mathcal{B} \subseteq \mathcal{T}$ that contains a neighbourhood base for $\mathcal{T}$ at each $x \in \mathcal{X}$. We say $\mathcal{X}$ is *second countable* if it has a countable base.

REFERENCES

ALEXANDROFF, P. (1924). Über die Metrisation der im Kleinen kompakten topologischen Räume. *Mathematische Annalen* **92** 294–301.

BERNSTEIN, S. (1914). Sur la meilleure approximation de $|x|$ par des polynomes de degrés donnés. *Acta Math.* **37** 1–57.

BERRETT, T. B. and SAMWORTH, R. J. (2023). Optimal nonparametric testing of Missing Completely At Random, and its connections to compatibility. *Submitted*.

BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

CAI, T. T. and LOW, M. G. (2011). Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *Ann. Statist.* **39** 1012–1041.

DÜMBGEN, L., SAMWORTH, R. J. and WELLNER, J. A. (2021). Bounding distributional errors via density ratios. *Bernoulli* **27** 818–852.

FOLLAND, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons.

GALE, D. (1957). A theorem on flows in networks. *Pacific J. Math* **7** 1073–1082.

ISII, K. (1964). Inequalities of the types of Chebyshev and Cramér-Rao and mathematical programming. *Ann. Inst. Statist. Math.* **16** 277–293.

JIAO, J., HAN, Y. and WEISSMAN, T. (2018). Minimax estimation of the L_1 distance. *IEEE Trans. Inf. Theory* **64** 6672–6706.

KELLERER, H. G. (1984). Duality theorems for marginal problems. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **67** 399–432.

MATOUSEK, J. and GÄRTNER, B. (2007). *Understanding and Using Linear Programming*. Springer Science & Business Media.

WU, Y. and YANG, P. (2016). Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Trans. Inf. Theory* **62** 3702–3720.