A Appendix: Proofs

A.1 Uniformity of rank

Throughout this appendix, let $\mathcal{T}$ be a non-empty finite or countably infinite set, let $\prec$ be a total order on $\mathcal{T}$ (of any order type), and let $\mathbf{p}$ and $\mathbf{q}$ each be a probability distribution on $\mathcal{T}$. For $n \in \mathbb{N}$, let $[n]$ denote the set $\{0, 1, 2, \ldots, n-1\}$.

Given a positive integer m, define the following random variables:

$$X_0 \sim \mathbf{q} \tag{13}$$

$$U_0 \sim \mathsf{Uniform}(0,1) \tag{14}$$

$$X_1, X_2, \ldots, X_m \sim^{\text{iid}} \mathbf{p} \tag{15}$$

$$U_1, U_2, \ldots, U_m \sim^{\text{iid}} \mathsf{Uniform}(0,1) \tag{16}$$

$$R = \textstyle\sum_{j=1}^{m} \mathbb{I}\left[X_j \prec X_0\right] + \mathbb{I}\left[X_j = X_0, U_j < U_0\right]. \tag{17}$$

Our first main result is the following, which establishes necessary and sufficient conditions for uniformity of the rank statistic.

Theorem A.1 (Theorem 3.1 in the main text)**.** *We have $\mathbf{p} = \mathbf{q}$ if and only if for all $m \geq 1$, the rank statistic R is uniformly distributed on $[m+1] \coloneqq \{0, 1, \ldots, m\}$.*

Before proving Theorem A.1, we state and prove several lemmas. We begin by showing that an i.i.d. sequence yields a uniform rank distribution.

Lemma A.2. *Let $T_0, T_1, \ldots, T_m$ be an i.i.d. sequence of random variables. If $\Pr\{T_i = T_j\} = 0$ for all distinct i and j, then the rank statistics $S_i \coloneqq \sum_{j=0}^{m} \mathbb{I}[T_j \prec T_i]$ for $0 \leq i \leq m$ are each uniformly distributed on $[m+1]$.*

Proof. Since $T_0, T_1, \ldots, T_m$ is i.i.d., it is a finitely exchangeable sequence, and so the rank statistics $S_0, \ldots, S_m$ are identically (but not independently) distributed.

Fix an arbitrary $k \in [m+1]$. Then $\Pr\{S_i = k\} = \Pr\{S_j = k\}$ for all $i, j \in [m+1]$. By hypothesis, $\Pr\{T_i = T_j\} = 0$ for distinct i and j. Therefore the rank statistics are almost surely distinct, and the events $\{S_i = j\}$ (for $0 \leq i \leq m$) are mutually exclusive and exhaustive. Since these events partition the outcome space, their probabilities sum to 1, and so $\Pr\{S_i = k\} = 1/(m+1)$ for all $i \in [m+1]$.

Because k was arbitrary, S_i is uniformly distributed on $[m+1]$ for all $i \in [m+1]$. ☐

We will also use the following result about convergence of discrete uniform variables to a continuous uniform random variable.

Lemma A.3. *Let $(V_m)_{m \geq 1}$ be a sequence of discrete random variables such that V_m is uniformly distributed on $\{0, 1/m, 2/m, \ldots, 1\}$, and let U be a continuous random variable uniformly distributed on the interval $[0,1]$. Then $(V_m)_{m\geq 1}$ converges in distribution to U, i.e.,*

$$\lim_{m\to\infty} \Pr\{V_m < u\} = \Pr\{U < u\} = u. \tag{18}$$

for all $u \in [0,1]$.

Furthermore, the convergence (18) is uniform in u.

Proof. Let $\epsilon > 0$. The distribution function F_m of V_m is given by

$$F_m(u) = \begin{cases} 1/(m+1) & u \in [0, 1/m) \\ 2/(m+1) & u \in [1/m, 2/m) \\ \cdots & \\ (a+1)/(m+1) & u \in [a/m, (a+1)/m) \\ \cdots & \\ m/(m+1) & u \in [(m-1)/m, 1) \\ 1 & u = 1. \end{cases}$$

Observe that for $0 \leq a < m$, the value $F_m(u)$ lies in the interval $[a/m, (a+1)/m)$ since we have that $(a/m) < (a+1)/(m+1) < (a+1)/m$. Since u is also in this interval, $|F_m(u) - u| \leq (a+1)/m - a/m = 1/m < \epsilon$ whenever $m > 1/\epsilon$, for all u. ☐

The following intermediate value lemma for step functions on the rationals is straightforward. It makes use of sums defined over subsets of the rationals, which are well-defined, as we discuss in the next remark.

Lemma A.4. *Let $p\colon (\mathbb{Q} \cap [0,1]) \to [0,1]$ be a function satisfying $p(0) = 0$ and $\sum_{x \in \mathbb{Q}\cap[0,1]} p(x) = 1$. Then for each $\delta \in (0,1)$, there is some $w \in \mathbb{Q} \cap [0,1]$ such that*

$$\sum_{x \in \mathbb{Q}\cap(0,w)} p(x) \leq \delta \leq \sum_{x \in \mathbb{Q}\cap(0,w]} p(x).$$

Remark A.5. The infinite sums in Lemma A.4 taken over a subset of the rationals can be formally defined as follows: Consider an arbitrary enumeration $\{q_1, q_2, \ldots, q_n, \ldots\}$ of $\mathbb{Q} \cap [0,1]$, and define the summation over the integer-valued index $n \geq 1$. Since the series consists of positive terms, it converges absolutely, and so all rearrangements of the enumeration converge to the same sum (see, e.g., [27, Theorem 3.55]).

One can show that the Cauchy criterion holds in this setting. Namely, suppose that a sum $\sum_{a<x<c} p(x)$ of non-negative terms converges. Then for all $\epsilon > 0$ there is some rational $b \in (a,c)$ such that $\sum_{a<x\leq b} p(x) < \epsilon$.

We now prove both directions of Theorem A.1.

Proof of Theorem A.1. Because $\mathcal{T}$ is countable, by a standard back-and-forth argument the total order $(\mathcal{T}, \prec)$ is isomorphic to $(B, <)$ for some subset $B \subseteq \mathbb{Q} \cap (0,1)$. Without loss of generality, we may therefore take $\mathcal{T}$ to be $\mathbb{Q} \cap [0,1]$ and assume that $\mathbf{p}(0) = \mathbf{p}(1) = 0$.

Consider the unit square $[0,1]^2$ equipped with the dictionary order $\lhd_{\mathrm{d}}$. This is a total order with the least upper bound property. For each $i \in [m+1]$, define $T_i \coloneqq (X_i, U_i)$, which takes values in $[0,1]^2$, and observe that the rank R in Eq. (6) of Theorem A.1 is equivalent to the rank $\sum_{i=0}^{m} \mathbb{I}[T_i \lhd_{\mathrm{d}} T_0]$ of T_0 taken according to the dictionary order.

(Necessity) Suppose $\mathbf{p} = \mathbf{q}$. Then $T_0, \dots, T_m$ are independent and identically distributed. Since $U_0, \dots, U_m$ are continuous random variables, we have $\Pr\{T_i = T_j\} = 0$ for all $i \neq j$. Apply Lemma A.2.

(Sufficiency) Suppose that for all $m > 0$, we have that the rank R is uniformly distributed on $\{0, 1, 2, \dots, m\}$. We begin the proof by first constructing a distribution function $F_{\mathbf{p}}$ on the unit square and then establishing several of its properties. First let $\tilde{\mathbf{p}} \colon [0,1] \to [0,1]$ be the "left-closed right-open" cumulative distribution function of $\mathbf{p}$, defined by

$$\tilde{\mathbf{p}}(x) \coloneqq \sum_{y \in \mathbb{Q} \cap [0,x)} \mathbf{p}(y)$$

for $x \in [0,1]$. Define $\mathbf{p}'$ to be the probability measure on $[0,1]$ that is equal to $\mathbf{p}$ on subsets of $\mathbb{Q} \cap [0,1]$ and is null elsewhere, and define the distribution function $F_{\mathbf{p}} \colon [0,1]^2 \to [0,1]$ on S by

$$F_{\mathbf{p}}(x,u) \coloneqq \tilde{\mathbf{p}}(x) + u\mathbf{p}'(x)$$

for $(x,u) \in [0,1]^2$. To establish that $F_{\mathbf{p}}$ is a valid distribution function, we show that its range is $[0,1]$; it is monotonically non-decreasing in each of its variables; and it is right-continuous in each of its variables.

It is immediate that $F_{\mathbf{p}}(0,0) = 0$ and $F_{\mathbf{p}}(1,1) = 1$. Furthermore, To establish that $F_{\mathbf{p}}$ is monotonically non-decreasing, put $x < y$ and $u < v$ and observe that

$$\begin{aligned} F_{\mathbf{p}}(x,u) &= \tilde{\mathbf{p}}(x) + u\mathbf{p}'(x) \\ &\leq \tilde{\mathbf{p}}(x) + \mathbf{p}'(x) \\ &\leq \textstyle\sum_{z \in \mathbb{Q} \cap [0,y)} \mathbf{p}'(z) \\ &= \tilde{\mathbf{p}}(y) \\ &\leq F_{\mathbf{p}}(y,u) \end{aligned}$$

and

$$\begin{aligned} F_{\mathbf{p}}(x,u) &= \tilde{\mathbf{p}}(x) + u\mathbf{p}'(x) \\ &\leq \tilde{\mathbf{p}}(x) + v\mathbf{p}'(x) \\ &= F_{\mathbf{p}}(x,v). \end{aligned}$$

We now establish right-continuity. For fixed x, $F_{\mathbf{p}}(x,u)$ is a linear function of u and so continuity is immediate. For fixed u, we have shown that $F_{\mathbf{p}}(x,u)$ is non-decreasing so it is sufficient to show that for any x and for any $\epsilon > 0$ there exists $x' > x$ such that

$$\begin{aligned} \epsilon &> F(x',u) - F(x,u) \\ &= \tilde{\mathbf{p}}(x') + u\mathbf{p}'(x') - \tilde{\mathbf{p}}(x) - u\mathbf{p}(x) \\ &= \tilde{\mathbf{p}}(x') + u\mathbf{p}'(x') - \tilde{\mathbf{p}}(x) - u\mathbf{p}(x) \\ &= \sum_{y \in \mathbb{Q} \cap [x,x']} \mathbf{p}(y), \end{aligned}$$

which is immediate from the Cauchy criterion.

Finally, we note that Lemma A.4 and the continuity of $F_{\mathbf{p}}$ in u together imply that $F_{\mathbf{p}}$ obtains all intermediate values, i.e., for any $\delta \in [0,1]$ there is some (x,u) such that $F(x,u) = \delta$.

Next define the inverse $F_{\mathbf{p}}^{-1} \colon [0,1] \to [0,1]^2$ by

$$F_{\mathbf{p}}^{-1}(s) \coloneqq \inf\{(x,u) \mid F_{\mathbf{p}}(x,u) = s\} \tag{19}$$

for $s \in [0,1]$, where the infimum is taken with the respect to the dictionary order $\lhd_{\mathrm{d}}$. The set in Eq (19) is non-empty since $F_{\mathbf{p}}$ obtains all values in $[0,1]$. Moreover, $F_{\mathbf{p}}^{-1}(s) \in [0,1]^2$ since $\lhd_{\mathrm{d}}$ has the least upper bound property. (This "generalized" inverse is used since $F_{\mathbf{p}}$ is one-to-one only under the stronger assumption that $\mathbf{p}(x) > 0$ for all $x \in \mathbb{Q} \cap (0,1)$.) Analogously define $F_{\mathbf{q}}$ in terms of $\mathbf{q}$.

Now define the rank function

$$r(a_0, \{a_1, \dots, a_m\}) \coloneqq \sum_{i=0}^{m} \mathbb{I}[a_i < a_0]$$

and note that $R \equiv r(T_0, \{T_1, \dots, T_m\})$. By the hypothesis, $r(T_0, \{T_1, \dots, T_m\})/m$ is uniformly distributed on $\{0, 1/m, 2/m, \dots, 1\}$ for all $m > 0$. Applying Lemma A.3 gives

$$\begin{aligned} \lim_{m \to \infty} \Pr\left\{\frac{1}{m}\tilde{r}(T_0, \{T_1, \dots, T_m\}) < s\right\} \\ = \Pr\{U_0 < s\} \\ = s. \end{aligned} \tag{20}$$

for $s \in [0,1]$.

For any $t \in [0,1]$ and $m \geq 1$, the random variable $\hat{F}_{\mathbf{p}}^{m}(t) \coloneqq \tilde{r}(t, \{T_1, \dots, T_m\})/m$ is the empirical distribution of $F_{\mathbf{p}}$. Therefore, by the Glivenko–Cantelli theorem for empirical distribution functions on k-dimensional Euclidean space [9, Corollary of Theorem 4], the sequence of random variables $(\hat{F}_{\mathbf{p}}^{m}(t))_{m \geq 1}$ converges a.s. to the real number $F_{\mathbf{p}}(t)$ uniformly in t, Hence the sequence $(\hat{F}_{\mathbf{p}}^{m}(T_0))_{m \geq 1}$ converges a.s. to the random variable $\hat{F}_{\mathbf{p}}(T_0)$, so that for any $s \in [0,1]$,

$$\begin{aligned}
\lim_{m\to\infty} \Pr\left\{\frac{1}{m}\tilde{r}(T_0, \{T_1,\dots,T_m\}) < s\right\} \\
&= \lim_{m\to\infty} \Pr\left\{\hat{F}^m_{\mathbf{p}}(T_0) < s\right\} && (21)\\
&= \Pr\left\{F_{\mathbf{p}}(T_0) < s\right\} && (22)\\
&= \Pr\left\{T_0 \lhd_{\mathrm{d}} F^{-1}_{\mathbf{p}}(s)\right\} && (23)\\
&= F_{\mathbf{q}}(F^{-1}_{\mathbf{p}}(s)). && (24)
\end{aligned}$$

The interchange of the limit and the probability in Eq. (22) follows from the bounded convergence theorem, since $\hat{F}^m_{\mathbf{p}}(T_0) \to F_{\mathbf{p}}(T_0)$ a.s. and for all $m \geq 1$ we have $|\hat{F}^m_{\mathbf{p}}(T_0)| \leq 1$ a.s.

Combining Eq. (20) and Eq. (24), we see that

$$F_{\mathbf{q}}(F^{-1}_{\mathbf{p}}(s)) = s \implies F^{-1}_{\mathbf{p}}(s) = F^{-1}_{\mathbf{q}}(s),$$

for $s \in [0,1]$. Since $0 \leq F_{\mathbf{p}}(x,u) \leq 1$, for each $(x,u) \in [0,1]^2$ we have

$$\begin{aligned}
F^{-1}_{\mathbf{q}}(F_{\mathbf{p}}(x,u)) &= F^{-1}_{\mathbf{p}}(F_{\mathbf{p}}(x,u))\\
&= F^{-1}_{\mathbf{q}}(F_{\mathbf{q}}(x,u))\\
&= (x,u).
\end{aligned}$$

It follows that $F_{\mathbf{p}}(x,u) = F_{\mathbf{q}}(x,u)$ for all $(x,u) \in [0,1]^2$. Fixing $u = 0$, we obtain

$$\tilde{\mathbf{p}}(x) = F_{\mathbf{p}}(x,0) = F_{\mathbf{q}}(x,0) = \tilde{\mathbf{q}}(x) \qquad (25)$$

for $x \in [0,1]$.

Assume, towards a contradiction, that $\mathbf{p} \neq \mathbf{q}$. Let a be any rational such that $\mathbf{p}(a) \neq \mathbf{q}(a)$, and suppose without loss of generality that $\mathbf{q}(a) < \mathbf{p}(a)$. By the Cauchy criterion (Remark A.4), there exists some $b > a$ such that

$$\sum_{a<x<b} \mathbf{q}(x) < \mathbf{p}(a) - \mathbf{q}(a).$$

Then we have

$$\begin{aligned}
\tilde{\mathbf{q}}(b) &= \tilde{\mathbf{q}}(a) + \mathbf{q}(a) + \sum_{x\in\mathbb{Q}\cap(a,b)} \mathbf{q}(x)\\
&= \tilde{\mathbf{p}}(a) + \mathbf{q}(a) + \sum_{x\in\mathbb{Q}\cap(a,b)} \mathbf{q}(x)\\
&< \tilde{\mathbf{p}}(a) + \mathbf{q}(a) + (\mathbf{p}(a) - \mathbf{q}(a))\\
&= \tilde{\mathbf{p}}(a) + \mathbf{p}(a)\\
&\leq \tilde{\mathbf{p}}(b),
\end{aligned}$$

and so $\tilde{\mathbf{p}} \neq \tilde{\mathbf{q}}$, contradicting Eq. (25). □

The following corollary is an immediate consequence.

Corollary A.6 (Corollary 3.3 in the main text)**.** *If* $\mathbf{p} \neq \mathbf{q}$, *then there is some* m *such that* R *is not uniformly distributed on* $[m+1]$.

The next theorem strengthens Corollary A.6 by showing that R is non-uniform for all but finitely many m.

Theorem A.7 (Theorem 3.4 in the main text)**.** *If* $\mathbf{p} \neq \mathbf{q}$, *then there is some* $M \geq 1$ *such that for all* $m \geq M$, *the rank* R *is not uniformly distributed on* $[m+1]$.

Before proving Theorem A.7, we show the following lemma.

Lemma A.8. *Suppose* $Z_1, \dots, Z_{m+1}$ *is a finitely exchangeable sequence of Bernoulli random variables. If*

$$S_m \coloneqq \sum_{i=1}^{m} Z_i$$

is not uniformly distributed on $[m+1]$, *then*

$$S_{m+1} \coloneqq \sum_{i=1}^{m+1} Z_i$$

is not uniformly distributed on $[m+2]$.

Proof. By finite exchangeability, there is some $r \in [0,1]$ such that the distribution of every Z_i is $\mathsf{Bernoulli}(r)$. There are two cases.

Case 1: $r \neq 1/2$. For any $\ell \geq 1$, we have

$$\mathbb{E}[S_\ell] = \mathbb{E}\left[\sum_{i=1}^{\ell} Z_i\right] = \sum_{i=1}^{\ell} \mathbb{E}[Z_i] = \ell r \neq r/2 = \mathbb{E}[U_\ell],$$

and so S_ℓ is not uniformly distributed on $[\ell+1]$. In particular, this holds for ℓ equal to either m or $m+1$, and so both the hypothesis and conclusion are true.

Case 2: $r = 1/2$. We prove the contrapositive. Suppose that S_{m+1} is uniformly distributed on $[m+1]$.

Assume S_{m+1} is uniform and fix $k \in [m+1]$. By total probability, we have

$$\begin{aligned}
\Pr\{S_m = k\} = \Pr\{S_m = k \text{ and } Z_{m+1} = 0\}\\
+ \Pr\{S_m = k \text{ and } Z_{m+1} = 1\}.
\end{aligned} \qquad (26)$$

We consider the two events on the right-hand side of Eq. (26) separately.

First, the event $\{S_m = k\} \cap \{Z_{m+1} = 0\}$ is the union over all $\binom{m}{k}$ assignments of $(Z_1, \dots, Z_m)$ that have exactly k ones and $Z_{m+1} = 0$. All such assignments are disjoint events. Define the event

$$\begin{aligned}
A \coloneqq \{Z_1 = \cdots = Z_k = 1\\
\text{and } Z_{k+1} = \cdots = Z_m = Z_{m+1} = 0\}.
\end{aligned}$$

By finite exchangeability, each assignment has probability $\Pr\{A\}$, and so

$$\Pr\{S_m = k \text{ and } Z_{m+1} = 0\} = \binom{m}{k} \Pr\{A\}. \quad (27)$$

Now, observe that the event $\{S_{m+1} = k\}$ is the union of all $\binom{m+1}{k}$ assignments of $(Z_1, \dots, Z_{m+1})$ that have exactly k ones. All the assignments are disjoint events and each has probability $\Pr\{A\}$, and so

$$\begin{aligned}\Pr\{S_{m+1} = k\} &= \binom{m+1}{k} \Pr\{A\} \\ &= \frac{1}{m+2}.\end{aligned} \quad (28)$$

Second, the event $\{S_m = k\} \cap \{Z_{m+1} = 1\}$ is the union over all $\binom{m}{k}$ assignments of $(Z_1, \dots, Z_m)$ that have exactly k ones and also $Z_{m+1} = 1$. All such assignments are disjoint events. Define the event

$$\begin{aligned}B \coloneqq \{&Z_1 = \dots = Z_k = Z_{m+1} = 1 \\ &\text{and } Z_{k+1} = \dots = Z_m = 0\}.\end{aligned}$$

Again by finite exchangeability, each assignment has probability $\Pr\{B\}$, and so

$$\Pr\{S_m = k \text{ and } Z_{m+1} = 1\} = \binom{m}{k} \Pr\{B\}. \quad (29)$$

Likewise, observe that the event $\{S_{m+1} = k+1\}$ is the union of all $\binom{m+1}{k+1}$ assignments of $(Z_1, \dots, Z_{m+1})$ that have exactly $k+1$ ones. All the assignments are disjoint events and each has probability $\Pr\{B\}$, and so

$$\begin{aligned}\Pr\{S_{m+1} = k+1\} &= \binom{m+1}{k+1} \Pr\{B\} \\ &= \frac{1}{m+2}.\end{aligned} \quad (30)$$

We now take Eq. (26), divide by $1/(m+2)$, and replace terms using Eqs. (27), (28), (29), and (30):

$$\begin{aligned}\frac{\Pr\{S_m = k\}}{1/(m+2)} &= \frac{\Pr\{S_m = k \text{ and } Z_{m+1} = 0\}}{1/(m+2)} \\ &\quad + \frac{\Pr\{S_m = k \text{ and } Z_{m+1} = 1\}}{1/(m+2)} \\ &= \frac{\binom{m}{k}\Pr\{A\}}{\binom{m+1}{k}\Pr\{A\}} + \frac{\binom{m}{k}\Pr\{B\}}{\binom{m+1}{k+1}\Pr\{B\}} \\ &= \frac{m!}{k!(m-k)!}\frac{k!(m+1-k)!}{(m+1)!} \\ &\quad + \frac{m!}{k!(m-k)!}\frac{(k+1)!(m+1-(k+1))!}{(m+1)!} \\ &= \frac{m+1-k}{m+1} + \frac{k+1}{m+1} \\ &= \frac{m+2}{m+1} \\ &= \frac{1/(m+1)}{1/(m+2)},\end{aligned}$$

and so we conclude that $\Pr\{S_m = k\} = 1/(m+1)$. □

We are now ready to prove Theorem A.7.

Proof of Theorem A.7. Suppose $\mathbf{p} \neq \mathbf{q}$. By Corollary A.6, there is some $M \geq 1$ such that the rank statistic $R = \sum_{i=1}^{M} \mathbb{I}[T_i \prec T_0]$ for $m = M$ is non-uniform over $[M+1]$. Observe that the rank statistic for $m = M+1$ is given by $\sum_{i=1}^{M+1} \mathbb{I}[T_i \prec T_0]$.

Now, each indicator $Z_i \coloneqq \mathbb{I}[T_i \prec T_0]$ is a Bernoulli random variable, and they are identically distributed since $(T_1, \dots, T_{M+1})$ is an i.i.d. sequence. Furthermore the sequence $(Z_1, \dots, Z_{M+1})$ is finitely exchangeable since the Z_i are conditionally independent given T_0. Then the sequence of indicators $(\mathbb{I}[T_1 \prec T_0], \mathbb{I}[T_2 \prec T_0], \dots, \mathbb{I}[T_{M+1} \prec T_0])$ satisfy the hypothesis of Lemma A.8, and so the rank statistic for $M+1$ is non-uniform. By induction, the rank statistic is non-uniform for all $m \geq M$. □

In fact, unless $\mathbf{p}$ and $\mathbf{q}$ satisfy an adversarial symmetry relationship under the selected ordering $\prec$, the rank is non-uniform for *any* choice of $m \geq 1$. Let $\lhd$ denote the lexicographic order on $\mathcal{T} \times [0,1]$ induced by $(\mathcal{T}, \prec)$ and $([0,1], <)$.

Corollary A.9 (Corollary 3.5 in the main text)**.** *Suppose* $\Pr\{(X, U_1) \lhd (Y, U_0)\} \neq 1/2$ *for* $Y \sim \mathbf{q}$, $X \sim \mathbf{p}$, *and* $U_0, U_1 \sim^{\text{iid}} \mathsf{Uniform}(0,1)$. *Then for all* $m \geq 1$, *the rank* R *is not uniformly distributed on* $[m+1]$.

Proof. If $\Pr\{(X, U_1) \lhd (Y, U_0)\} \neq 1/2$ then R is non-uniform for $m = 1$. The conclusion follows by Theorem A.7. □

A.2 An ordering that witnesses $\mathbf{p} \neq \mathbf{q}$ for $m = 1$

We now describe an ordering $\prec$ for which, when $m = 1$, we have $\Pr\{R = 0\} > 1/2$.

Define

$$A \coloneqq \{x \in \mathcal{T} \mid \mathbf{q}(x) > \mathbf{p}(x)\}$$

to be the set of all elements of $\mathcal{T}$ that have a greater probability according to $\mathbf{q}$ than according to $\mathbf{p}$, and let A^c denote its complement. Let $\mathbf{h}_{\mathbf{p},\mathbf{q}}$ be the signed measure given by the difference $\mathbf{h}_{\mathbf{p},\mathbf{q}}(x) \coloneqq \mathbf{q}(x) - \mathbf{p}(x)$

between $\mathbf{q}$ and $\mathbf{p}$; for the rest of this subsection, we denote this simply by $\mathbf{h}$. Let $\prec$ be any total order on $\mathcal{T}$ satisfying

- if $\mathbf{h}(x) > \mathbf{h}(x')$ then $x \prec x'$; and
- if $\mathbf{h}(x) < \mathbf{h}(x')$ then $x \succ x'$.

The linear ordering $\prec$ may be defined arbitrarily for all pairs x and x' which satisfy $\mathbf{h}(x) = \mathbf{h}(x')$. As an immediate consequence, $x \prec x'$ whenever $x \in A$ and $x' \in A^c$. Intuitively, the ordering is designed to ensure that elements $x \in A$ are "small", and are ordered by decreasing value of $\mathbf{q}(x) - \mathbf{p}(x)$ (with ties broken arbitrarily); elements $x \in A^c$ are "large" and are ordered by increasing value of $\mathbf{p}(x) - \mathbf{q}(x)$ (again, with ties broken arbitrarily). The smallest element in $\mathcal{T}$ maximizes $\mathbf{q}(x) - \mathbf{p}(x)$ and the largest element in $\mathcal{T}$ maximizes $\mathbf{p}(x) - \mathbf{q}(x)$.

We first establish some easy lemmas.

Lemma A.10. *$A = \emptyset$ if and only if $\mathbf{p} = \mathbf{q}$.*

Proof. Immediate. □

Lemma A.11.

$$\sum_{x \in A} [\mathbf{q}(x) - \mathbf{p}(x)] = \sum_{x \in A^c} [\mathbf{p}(x) - \mathbf{q}(x)].$$

Proof. We have

$$\begin{aligned}&\sum_{x \in A}[\mathbf{q}(x) - \mathbf{p}(x)] - \sum_{x \in A^c} [\mathbf{p}(x) - \mathbf{q}(x)] \\ &\quad = \sum_{x \in \mathcal{T}} \mathbf{q}(x) - \sum_{x \in \mathcal{T}} \mathbf{p}(x) = 0,\end{aligned}$$

as desired. □

Given a probability distribution $\mathbf{r}$, define its cumulative distribution function $\tilde{\mathbf{r}}$ by $\tilde{\mathbf{r}}(x) \coloneqq \sum_{y \prec x} \mathbf{r}(y)$.

Lemma A.12. *$\tilde{\mathbf{q}}(x) > \tilde{\mathbf{p}}(x)$ for all $x \in \mathcal{T}$.*

Proof. Let $\mathcal{T}_x \coloneqq \{y \in \mathcal{T} \mid y \prec x\}$. If $x \in A$ then $\mathcal{T}_x \subseteq A$, and so

$$\tilde{\mathbf{q}}(x) - \tilde{\mathbf{p}}(x) = \sum_{y \in \mathcal{T}_x} [\mathbf{q}(y) - \mathbf{p}(y)] > 0,$$

since all terms in the sum are positive.

Otherwise, $y \in A$ for all $y \prec x$, and so $A \subseteq \mathcal{T}_x$. Let $A^c_x \coloneqq \{y \in A^c \mid y \prec x\}$. Then

$$\begin{aligned}&\tilde{\mathbf{q}}(x) - \tilde{\mathbf{p}}(x) \\ &= \sum_{y \prec x}[\mathbf{q}(y) - \mathbf{p}(y)] \\ &= \sum_{y \in A}[\mathbf{q}(y) - \mathbf{p}(y)] + \sum_{y \in A^c_x} [\mathbf{q}(y) - \mathbf{p}(y)] \\ &= \sum_{y \in A_x} [\mathbf{q}(y) - \mathbf{p}(y)] - \sum_{y \in A^c_x} [\mathbf{p}(y) - \mathbf{q}(y)] \\ &> \sum_{y \in A_x} [\mathbf{q}(y) - \mathbf{p}(y)] - \sum_{y \in A^c} [\mathbf{p}(y) - \mathbf{q}(y)] \\ &= 0,\end{aligned}$$

establishing the lemma. □

We now analyze $\Pr\{R = 0\}$ in the case where $m = 1$. In this case, we may drop some subscripts and write Y in place of X_1, so that our setting reduces to the following random variables:

$$\begin{aligned} X_{\mathbf{p}} &\sim \mathbf{p} \\ Y_{\mathbf{q}} &\sim \mathbf{q} \\ R_{\mathbf{p},\mathbf{q}} \mid X_{\mathbf{p}}, Y_{\mathbf{q}} &\sim \begin{cases} 0 & \text{if } X_{\mathbf{p}} \succ Y_{\mathbf{q}}, \\ 1 & \text{if } X_{\mathbf{p}} \prec Y_{\mathbf{q}}, \\ \text{Bernoulli}(1/2) & \text{if } X_{\mathbf{p}} = Y_{\mathbf{q}}. \end{cases}\end{aligned}$$

(We have indicated $\mathbf{p}$ and $\mathbf{q}$ in the subscripts, for use in the next subsection.)

In other words, the procedure samples $X_{\mathbf{p}} \sim \mathbf{p}$ and $Y_{\mathbf{q}} \sim \mathbf{q}$ independently. Given these values, it then sets $R_{\mathbf{p},\mathbf{q}}$ to be 0 if $X_{\mathbf{p}} \succ Y_{\mathbf{q}}$, to be 1 if $X_{\mathbf{p}} \prec Y_{\mathbf{q}}$, and the outcome of an independent fair coin flip otherwise.

For the rest of this subsection, we will refer to these random variables simply as X, Y, and R, though later on we will need them for several choices of distributions $\mathbf{p}$ and $\mathbf{q}$ (and accordingly will retain the subscripts).

We now prove the following theorem.

Theorem A.13 (Theorem 3.6 in the main text)**.** *If $\mathbf{p} \neq \mathbf{q}$, then for $m = 1$ and the ordering $\prec$ defined above, we have* $\Pr\{R = 0\} > 1/2$.

Proof. From total probability and independence of X and Y, we have

$$\begin{aligned}&\Pr\{R = 0\} \\ &= \sum_{x,y \in \mathcal{T}} \Pr\{R{=}0 \mid X{=}x, Y{=}y\} \Pr\{Y = y\} \Pr\{X = x\} \\ &= \sum_{x,y \in \mathcal{T}} \Pr\{R{=}0 \mid X{=}x, Y{=}y\}\, \mathbf{q}(y)\mathbf{p}(x) \\ &= \sum_{x \in \mathcal{T}} \Pr\{R{=}0 \mid X{=}x, Y{=}x\}\, \mathbf{q}(x)\mathbf{p}(x) \\ &\qquad + \sum_{y \prec x \in \mathcal{T}} \Pr\{R{=}0 \mid X{=}x, Y{=}y\}\, \mathbf{q}(y)\mathbf{p}(x)\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{x \prec y \in \mathcal{T}} \Pr\{R{=}0 \mid X{=}x, Y{=}y\} \mathbf{q}(y)\mathbf{p}(x) \\
&= \frac{1}{2}\sum_{x\in\mathcal{T}} \mathbf{q}(x)\mathbf{p}(x) + 1 \sum_{y \prec x \in \mathcal{T}} \mathbf{q}(y)\mathbf{p}(x) \\
&\quad + 0 \sum_{x \prec y \in \mathcal{T}} \mathbf{q}(y)\mathbf{p}(x) \\
&= \frac{1}{2}\sum_{x\in\mathcal{T}} \mathbf{p}(x)\mathbf{q}(x) + \sum_{x\in\mathcal{T}} \tilde{\mathbf{q}}(x)\mathbf{p}(x).
\end{aligned}
$$

An identical argument establishes that

$$
\Pr\{R{=}1\} = \frac{1}{2}\sum_{x\in\mathcal{T}} \mathbf{p}(x)\mathbf{q}(x) + \sum_{x\in\mathcal{T}} \tilde{\mathbf{p}}(x)\mathbf{q}(x).
$$

Since $\Pr\{R{=}0\} + \Pr\{R = 1\} = 1$, it suffices to establish that $\Pr\{R = 0\} > \Pr\{R = 1\}$. We have

$$
\begin{aligned}
&\Pr\{R = 0\} - \Pr\{R = 1\} \\
&= \sum_{x\in\mathcal{T}} \tilde{\mathbf{q}}(x)\mathbf{p}(x) - \sum_{x\in\mathcal{T}} \tilde{\mathbf{p}}(x)\mathbf{q}(x) \\
&> \sum_{x\in\mathcal{T}} \tilde{\mathbf{p}}(x)\mathbf{p}(x) - \sum_{x\in\mathcal{T}} \tilde{\mathbf{p}}(x)\mathbf{q}(x) \\
&= \sum_{x\in\mathcal{T}} \tilde{\mathbf{p}}(x)[\mathbf{p}(x) - \mathbf{q}(x)] \\
&= \sum_{x\in A^c} \tilde{\mathbf{p}}(x)[\mathbf{p}(x) - \mathbf{q}(x)] - \sum_{x\in A} \tilde{\mathbf{p}}(x)[\mathbf{q}(x) - \mathbf{p}(x)] \\
&\geq \sum_{x\in A^c} \big(\max_{y\in A} \tilde{\mathbf{p}}(y)\big)[\mathbf{p}(x) - \mathbf{q}(x)] \\
&\qquad - \sum_{x\in A} \tilde{\mathbf{p}}(x)[\mathbf{q}(x) - \mathbf{p}(x)] \\
&= \sum_{x\in A} \big(\max_{y\in A} \tilde{\mathbf{p}}(y)\big)[\mathbf{q}(x) - \mathbf{p}(x)] \\
&\qquad - \sum_{x\in A} \tilde{\mathbf{p}}(x)[\mathbf{q}(x) - \mathbf{p}(x)] \\
&= \sum_{x\in A} \big(\max_{y\in A} \tilde{\mathbf{p}}(y) - \tilde{\mathbf{p}}(x)\big)[\mathbf{q}(x) - \mathbf{p}(x)] \\
&> 0.
\end{aligned}
$$

The first inequality follows from Lemma A.12; the second inequality follows from monotonicity of $\tilde{\mathbf{p}}$; the second-to-last equality follows from Lemma A.11; and the final inequality follows from the fact that all terms in the sum are positive. □

A.3 A tighter bound in terms of $L_\infty(\mathbf{p}, \mathbf{q})$

We have just exhibited an ordering such that when $\mathbf{p} \neq \mathbf{q}$ and $m = 1$, we have $\Pr\{R = 0\} > 1/2$. We are now interested in obtaining a tighter lower bound on this probability in terms of the L_∞ distance between $\mathbf{p}$ and $\mathbf{q}$.

In this subsection and the following one, we assume that $\mathcal{T}$ is finite. We first note the following immediate lemma.

Lemma A.14. *Let $B, C \subseteq \mathcal{T}$. For all $\mathbf{p}, \mathbf{q}$ and all $\delta > 0$ there is an $\epsilon > 0$ such that for all distributions $\mathbf{p}'$ on $\mathcal{T}$ with $\sup_{x\in\mathcal{T}} |\mathbf{p}(x) - \mathbf{p}'(x)| < \epsilon$, we have*

$$
\begin{aligned}
&\big|\Pr(R_{\mathbf{p},\mathbf{q}} = 0 \mid X_{\mathbf{p}} \in B,\ Y_{\mathbf{q}} \in C) \\
&\quad - \Pr(R_{\mathbf{p}',\mathbf{q}} = 0 \mid X_{\mathbf{p}'} \in B,\ Y_{\mathbf{q}} \in C)\big| < \delta.
\end{aligned}
$$

Definition A.15. We say that $\mathbf{p}$ is **ϵ-discrete** (with respect to $\mathbf{q}$) if for all $a, b \in \mathcal{T}$ we have

$$
\big|\mathbf{h}_{\mathbf{p},\mathbf{q}}(a) - \mathbf{h}_{\mathbf{p},\mathbf{q}}(b)\big| \geq \epsilon.
$$

From Lemma A.14 we immediately obtain the following.

Lemma A.16. *For all $\mathbf{p}, \mathbf{q}$ and all $\delta > 0$ there is an $\epsilon > 0$ and an ϵ-discrete distribution $\mathbf{p}_\epsilon$ on $\mathcal{T}$ such that for all $B, C \subseteq \mathcal{T}$,*

$$
\begin{aligned}
&\big|\Pr(R_{\mathbf{p},\mathbf{q}} = 0 \mid X_{\mathbf{p}} \in B,\ Y_{\mathbf{q}} \in C) \\
&\quad - \Pr(R_{\mathbf{p}_\epsilon,\mathbf{q}} = 0 \mid X_{\mathbf{p}_\epsilon} \in B,\ Y_{\mathbf{q}} \in C)\big| < \delta.
\end{aligned}
$$

The next lemma will be crucial for proving our bound.

Lemma A.17. *Let $\mathbf{p}_0$ and $\mathbf{p}_1$ be probability measures on $\mathcal{T}$, and let $\lhd$ be a total order on $\mathcal{T}$ such that if $\mathbf{h}_{\mathbf{p}_0,\mathbf{q}}(x) > \mathbf{h}_{\mathbf{p}_0,\mathbf{q}}(x')$ then $x \lhd x'$ and if $\mathbf{h}_{\mathbf{p}_0,\mathbf{q}}(x) < \mathbf{h}_{\mathbf{p}_0,\mathbf{q}}(x')$ then $x \rhd x'$. Suppose that if $\mathbf{h}_{\mathbf{p}_0,\mathbf{p}_1}(x) > 0$ and $\mathbf{h}_{\mathbf{p}_0,\mathbf{p}_1}(y) \leq 0$, then $x \lhd y$. Then $\Pr(R_{\mathbf{p}_0,\mathbf{q}} = 0) \geq \Pr(R_{\mathbf{p}_1,\mathbf{q}} = 0)$.*

Proof. Note that

$$
\begin{aligned}
&\Pr(R_{\mathbf{p}_1,\mathbf{q}} = 0 \mid Y_{\mathbf{q}} = y) \\
&= \sum_{x \rhd y} \mathbf{p}_1(x) + \frac{1}{2}\mathbf{p}_1(y) \\
&= \sum_{x \rhd y} \mathbf{p}_0(x) + \mathbf{h}_{\mathbf{p}_0,\mathbf{p}_1}(x) + \frac{1}{2}[\mathbf{p}_0(y) + \mathbf{h}_{\mathbf{p}_0,\mathbf{p}_1}(y)] \\
&= \Pr(R_{\mathbf{p}_0,\mathbf{q}} = 0 \mid Y_{\mathbf{q}} = y) + \sum_{x \rhd y} \mathbf{h}_{\mathbf{p}_0,\mathbf{p}_1}(x) + \frac{1}{2}\mathbf{h}_{\mathbf{p}_0,\mathbf{p}_1}(y) \\
&= \Pr(R_{\mathbf{p}_0,\mathbf{q}} = 0 \mid Y_{\mathbf{q}} = y) - \sum_{x \lhd y} \mathbf{h}_{\mathbf{p}_0,\mathbf{p}_1}(x) - \frac{1}{2}\mathbf{h}_{\mathbf{p}_0,\mathbf{p}_1}(y),
\end{aligned}
$$

where the last equality holds because $\sum_{x\in\mathcal{T}} \mathbf{h}_{\mathbf{p}_0,\mathbf{p}_1}(x) = 0$. But by our assumption, we know that $\sum_{x \lhd y} \mathbf{h}_{\mathbf{p}_0,\mathbf{p}_1}(x) + \frac{1}{2}\mathbf{h}_{\mathbf{p}_0,\mathbf{p}_1}(y)$ is non-negative and so $\Pr(R_{\mathbf{p}_1,\mathbf{q}} = 0 \mid Y_{\mathbf{q}} = y) \leq \Pr(R_{\mathbf{p}_0,\mathbf{q}} = 0 \mid Y_{\mathbf{q}} = y)$, from which the result follows. □

We will now provide a lower bound on $\Pr(R_{\mathbf{p},\mathbf{q}} = 0)$.

Proposition A.18.

$$\Pr(R_{\mathbf{p},\mathbf{q}} = 0) \geq \frac{1}{2} + \frac{1}{2} \max_{x \in \mathcal{T}} \mathbf{h}_{\mathbf{p},\mathbf{q}}(x)^2. \quad (31)$$

Proof. Recall that $A \coloneqq \{x \in \mathcal{T} \mid \mathbf{q}(x) > \mathbf{p}(x)\}$. First note that by Lemma A.14, we may assume without loss of generality that $|A| = |\mathcal{T} \setminus A|$, by adding elements of mass arbitrarily close to 0. Let $k \coloneqq |A|$. Further, by Lemma A.16 we may assume without loss of generality that $\mathbf{p}, \mathbf{q}$ are an ϵ-discrete pair (for some fixed but small ϵ) with $|\mathcal{T}| \cdot \epsilon < L_\infty(\mathbf{p}, \mathbf{q})$. Let $(x_0^+, \dots, x_{k-1}^+)$ be the collection A listed in $\prec$-increasing order. Let $(x_0^-, \dots, x_{k-1}^-)$ be the collection $\mathcal{T} \setminus A$ listed in $\prec$-increasing order.

Let $\mathbf{p}^*$ be any probability measure such that

$$\mathbf{p}^*(x) = \begin{cases} \mathbf{p}(x) - e(\ell) & (x = x_\ell^-; e(\ell) \geq 0), \\ \mathbf{q}(x) - (k - \ell) \cdot \epsilon & (x = x_\ell^+; 0 \leq \ell < k - 1), \\ \mathbf{p}(x) & (x = x_0^+). \end{cases}$$

Note that for all $x, y \in \mathcal{T}$, we have $y \prec x$ if and only if $\mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(x) < \mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(y)$.

Now, for every $\ell < k - 1$ we have $\mathbf{h}_{\mathbf{p},\mathbf{q}}(x_\ell^+) \geq \ell \cdot \epsilon$ (as $\mathbf{p}, \mathbf{q}$ are an ϵ-discrete pair), and so we can always find such a $\mathbf{p}^*$. In particular the following are immediate.

(a) $x \prec y$ if and only if $\mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(x) > \mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(y)$,

(b) $\mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+) = \mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(x_0^+)$,

(c) if $\mathbf{h}_{\mathbf{p},\mathbf{q}^*}(x) > 0$ and $\mathbf{h}_{\mathbf{p},\mathbf{p}^*}(y) \leq 0$ then $x \prec y$, and

(d) $(\mathbf{p}, \mathbf{q}^*)$ is an ϵ-discrete pair.

Note that $\Pr(R_{\mathbf{p},\mathbf{q}} = 0) \geq \Pr(R_{\mathbf{p}^*,\mathbf{q}} = 0)$, by Lemma A.17 and (c). For simplicity, let $A_0 \coloneqq \{x_0^+\}$, $A_1 \coloneqq \{x_i^+\}_{1 \leq i \leq k-1}$ and $D \coloneqq \mathcal{T} \setminus A$.

We now condition on the value of $Y_\mathbf{q}$, in order to calculate $\Pr(R_{\mathbf{p}^*,\mathbf{q}} = 0)$.

Case 1: $Y_\mathbf{q} = x_i^-$. We have

$$\Pr(R_{\mathbf{p}^*,\mathbf{q}} = 0 \mid Y_\mathbf{q} = x_i^-) = \sum_{i<\ell<k} \mathbf{p}^*(x_\ell^-) + \frac{1}{2}\mathbf{p}^*(x_i^-).$$

Case 2: $Y_\mathbf{q} \in A_1$. We have

$$\Pr(R_{\mathbf{p}^*,\mathbf{q}} = 0 \mid Y_\mathbf{q} \in A_1) = \mathbf{p}^*(D) + \frac{1}{2}\mathbf{p}^*(A_1) + f_0(\epsilon),$$

where f_0 is a function satisfying $\lim_{\epsilon \to 0} f_0(\epsilon) = 0$.

Case 3: $Y_\mathbf{q} \in A_0$. We have

$$\Pr(R_{\mathbf{p}^*,\mathbf{q}} = 0 \mid Y_\mathbf{q} \in A_0) = \mathbf{p}^*(A_1) + \mathbf{p}^*(D) + \frac{1}{2}\mathbf{p}^*(A_0).$$

We may calculate these terms as follows:

$$\begin{aligned} \mathbf{p}^*(D) &= \mathbf{q}(D) + \mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+) + (k(k-1)/2)\epsilon, \\ \mathbf{p}^*(A_1) &= \mathbf{q}(A_1) - (k(k-1)/2)\epsilon, \\ \mathbf{p}^*(A_0) &= \mathbf{q}(A_0) - \mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+). \end{aligned}$$

Putting all of this together, we obtain

$$\begin{aligned} &\Pr(R_{\mathbf{p}^*,\mathbf{q}} = 0) \\ &= \sum_{i<k} \sum_{i<\ell<k} \mathbf{q}(x_i^-)\mathbf{p}^*(x_\ell^-) + \frac{1}{2}\sum_{i<k} \mathbf{q}(x_i^-)\mathbf{p}^*(x_i^-) \\ &\quad + \mathbf{q}(A_1)\mathbf{p}^*(D) + \frac{1}{2}\mathbf{q}(A_1)\mathbf{p}^*(A_1) + \mathbf{q}(A_1) f_0(\epsilon) \\ &\quad + \mathbf{q}(A_0)\mathbf{p}^*(A_1) + \mathbf{q}(A_0)\mathbf{p}^*(D) + \frac{1}{2}\mathbf{q}(A_0)\mathbf{p}^*(A_0) \\ &= \sum_{i<k} \sum_{i<\ell<k} \mathbf{q}(x_i^-)[\mathbf{q}(x_\ell^-) - \mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(x_\ell^-)] \\ &\quad + \frac{1}{2}\sum_{i<k} \mathbf{q}(x_i^-)[\mathbf{q}(x_i^-) - \mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(x_i^-)] \\ &\quad + \mathbf{q}(A_1)[\mathbf{q}(D) + \mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+)] + \frac{1}{2}\mathbf{q}(A_1)\mathbf{q}(A_1) \\ &\quad + \mathbf{q}(A_0)\mathbf{q}(A_1) + \mathbf{q}(A_0)[\mathbf{q}(D) + \mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+)] \\ &\quad + \frac{1}{2}\mathbf{q}(A_0)[\mathbf{q}(A_0) - \mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+)] + f_1(\epsilon) \\ &= \sum_{i<k} \sum_{i<\ell<k} \mathbf{q}(x_i^-)\mathbf{q}(x_\ell^-) + \frac{1}{2}\sum_{i<k} \mathbf{q}(x_i^-)\mathbf{q}(x_i^-) \\ &\quad + \mathbf{q}(A_1)\mathbf{q}(D) + \frac{1}{2}\mathbf{q}(A_1)\mathbf{q}(A_1) + \mathbf{q}(A_0)\mathbf{q}(A_1) \\ &\quad + \mathbf{q}(A_0)\mathbf{q}(D) + \frac{1}{2}\mathbf{q}(A_0)\mathbf{q}(A_0) \\ &\quad - \sum_{i<k} \sum_{i<\ell<k} \mathbf{q}(x_i^-)\mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(x_\ell^-) \\ &\quad - \frac{1}{2}\sum_{i<k} \mathbf{q}(x_i^-)\mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(x_i^-) + \mathbf{q}(A_1)\mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+) \\ &\quad + \mathbf{q}(A_0)\mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+) - \frac{1}{2}\mathbf{q}(A_0)\mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+) + f_1(\epsilon) \\ &= \sum_{i<k} \sum_{i<\ell<k} \mathbf{q}(x_i^-)\mathbf{q}(x_\ell^-) + \frac{1}{2}\sum_{i<k} \mathbf{q}(x_i^-)\mathbf{q}(x_i^-) \\ &\quad + \mathbf{q}(A_1)\mathbf{q}(D) + \frac{1}{2}\mathbf{q}(A_1)\mathbf{q}(A_1) + \mathbf{q}(A_0)\mathbf{q}(A_1) \\ &\quad + \mathbf{q}(A_0)\mathbf{q}(D) + \frac{1}{2}\mathbf{q}(A_0)\mathbf{q}(A_0) \\ &\quad - \sum_{i<k} \sum_{i<\ell<k} \mathbf{q}(x_i^-)\mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(x_\ell^-) \\ &\quad - \frac{1}{2}\sum_{i<k} \mathbf{q}(x_i^-)\mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(x_i^-) + \mathbf{q}(A_1)\mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+) \\ &\quad + \frac{1}{2}\mathbf{q}(A_0)\mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+) + f_1(\epsilon), \end{aligned}$$

where f_1 is a function satisfying $\lim_{\epsilon\to 0} f_1(\epsilon) = 0$.

We also have

$$\begin{aligned}\frac{1}{2} &= \Pr(R_{\mathbf{q},\mathbf{q}} = 0)\\ &= \sum_{i<k}\sum_{i<\ell<k} \mathbf{q}(x_i^-)\mathbf{q}(x_\ell^-) + \frac{1}{2}\sum_{i<k}\mathbf{q}(x_i^-)\mathbf{q}(x_i^-)\\ &\quad + \mathbf{q}(A_1)\mathbf{q}(D) + \frac{1}{2}\mathbf{q}(A_1)\mathbf{q}(A_1)\\ &\quad + \mathbf{q}(A_0)\mathbf{q}(A_1) + \mathbf{q}(A_0)\mathbf{q}(D) + \frac{1}{2}\mathbf{q}(A_0)\mathbf{q}(A_0).\end{aligned}$$

Putting these two equations together, we obtain

$$\begin{aligned}&\Pr(R_{\mathbf{p}^*,\mathbf{q}} = 0) - \frac{1}{2}\\ &= \Pr(R_{\mathbf{p}^*,\mathbf{q}} = 0) - \Pr(R_{\mathbf{q},\mathbf{q}} = 0)\\ &= -\sum_{i<k}\sum_{i<\ell<k} \mathbf{q}(x_i^-)\mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(x_\ell^-)\\ &\quad - \frac{1}{2}\sum_{i<k}\mathbf{q}(x_i^-)\mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(x_i^-) + \mathbf{q}(A_1)\mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+)\\ &\quad + \frac{1}{2}\mathbf{q}(A_0)\mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+) + f_1(\epsilon)\\ &\geq \frac{1}{2}\mathbf{q}(A_0)\mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+) + f_1(\epsilon),\end{aligned}$$

as $\mathbf{h}_{\mathbf{p}^*,\mathbf{q}}(x_\ell^-) \leq 0$ for all $\ell < k$ and $\mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+) \leq 0$.

But we know that

$$\mathbf{q}(A_0) = \mathbf{q}(x_0^+) = \mathbf{p}^*(x_0^+) + \mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+) \geq \mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+).$$

Therefore, as $\mathbf{h}_{\mathbf{p},\mathbf{q}}(x_0^+)$ is the maximal value of $\mathbf{h}_{\mathbf{p},\mathbf{q}}$, by taking the limit as $\epsilon \to 0$ we obtain

$$\Pr(R_{\mathbf{p}^*,\mathbf{q}} = 0) \geq \frac{1}{2} + \frac{1}{2}\max_{x\in\mathcal{T}} \mathbf{h}_{\mathbf{p},\mathbf{q}}(x)^2,$$

as desired. □

Finally, we arrive at the following theorem.

Theorem A.19. *Given probability measure $\mathbf{p}, \mathbf{q}$ on $\mathcal{T}$ there is a linear ordering $\sqsubset$ of $\mathcal{T}$ such that if $X_\mathbf{p}$ and $Y_\mathbf{q}$ are sampled independently from $\mathbf{p}$ and $\mathbf{q}$ respectively then*

$$\Pr(X_\mathbf{q} \sqsubset Y_\mathbf{p}) \geq \frac{1}{2} + \frac{1}{2}L_\infty(\mathbf{p},\mathbf{q})^2. \tag{32}$$

Proof. Note that

$$L_\infty(\mathbf{p},\mathbf{q}) = \max\{\max_{x\in\mathcal{T}} \mathbf{h}_{\mathbf{p},\mathbf{q}}(x), \max_{x\in\mathcal{T}} \mathbf{h}_{\mathbf{q},\mathbf{p}}(x)\}.$$

If $L_\infty(\mathbf{p},\mathbf{q}) = \max_{x\in\mathcal{T}} \mathbf{h}_{\mathbf{p},\mathbf{q}}(x)$, then the theorem follows from Proposition A.18 using the ordering $x \sqsubset y$ if and only if $\mathbf{h}_{\mathbf{p},\mathbf{q}}(x) > \mathbf{h}_{\mathbf{p},\mathbf{q}}(y)$.

If, however, $L_\infty(\mathbf{p},\mathbf{q}) = \max_{x\in\mathcal{T}} \mathbf{h}_{\mathbf{q},\mathbf{p}}(x)$, then the theorem follows from Proposition A.18 by interchanging $\mathbf{p}$ and $\mathbf{q}$, i.e., by using the ordering $x \sqsubset y$ if and only if $\mathbf{h}_{\mathbf{q},\mathbf{p}}(x) > \mathbf{h}_{\mathbf{q},\mathbf{p}}(y)$. □

A.4 Sample complexity

We now show how to amplify this result by repeated trials to obtain a bound on the sample complexity of the main algorithm for determining whether $\mathbf{p} = \mathbf{q}$.

Let $\sqsubset$ be the linear ordering defined in Theorem A.19.

Theorem A.20 (Theorem 3.7 in the main text)**.** *Given significance level $\alpha = 2\Phi(-c)$ for $c > 0$, the proposed test with ordering $\sqsubset$ and $m = 1$ achieves power $\beta \geq 1 - \Phi(-c)$ using*

$$n \approx 4c^2/L_\infty(\mathbf{p},\mathbf{q})^4 \tag{33}$$

samples from $\mathbf{q}$, where Φ is the cumulative distribution function of a standard normal.

Proof. Assume without loss of generality that the order $\sqsubset$ from Theorem A.19 is such that $L_\infty = \max_{s\in\mathcal{T}}(\mathbf{q}(x) - \mathbf{p}(x))$. Let $(Y_1,\dots,Y_n) \sim^{\text{iid}} \mathbf{q}$ be the n samples from $\mathbf{q}$. With $m = 1$, the testing procedure generates n samples $(X_1,\dots,X_n) \sim^{\text{iid}} \mathbf{p}$, and $2n$ uniform random variables $(U_1^Y,\dots,U_n^Y,U_1^X,\dots,U_n^X) \sim^{\text{iid}}$ $\mathsf{Uniform}(0,1)$ to break ties. Let $\lhd$ denote the lexicographic order on $\mathcal{T}\times[0,1]$ induced by $(\mathcal{T},\lhd)$ and $([0,1],<)$. Define $W_i \coloneqq \mathbb{I}\left[(Y_i,U_i^Y) \lhd (X_i,U_i^X)\right]$, for $1 \leq i \leq n$, to be the rank of the i-th observation from $\mathbf{q}$.

Under the null hypothesis H_0, each rank W_i has distribution $\mathsf{Bernoulli}(1/2)$ by Lemma A.2. Testing for uniformity of the ranks on $\{0,1\}$ is equivalent to testing whether a coin is unbiased given the i.i.d. flips $\{W_1,\dots,W_n\}$. Let $\hat{B} \coloneqq \sum_{i=1}^n (1-W_i)/n$ denote the empirical proportion of zeros. By the central limit theorem, for sufficiently large n, we have that $\hat{B}$ is approximately normally distributed with mean $1/2$ and standard deviation $1/(2\sqrt{n})$. For the given significance level $\alpha = 2\Phi(-c)$, we form the two-sided reject region $F = (-\infty,\gamma)\cup(\gamma,\infty)$, where the critical value γ satisfies

$$c = \frac{\gamma - 1/2}{1/(2\sqrt{n})} = 2\sqrt{n}(\gamma - 1/2). \tag{34}$$

Replacing n in Eq. (7), we obtain

$$\begin{aligned}\gamma &= 1/2 + c/(2\sqrt{n})\\ &= 1/2 + c/(2(2c/L_\infty(\mathbf{p},\mathbf{q})^2))\\ &= 1/2 + L_\infty(\mathbf{p},\mathbf{q})^2/4.\end{aligned} \tag{35}$$

This construction ensures that $\Pr\{\mathsf{reject} \mid \mathsf{H}_0\} = \alpha$.

We now show that the test with this rejection region has power $\beta \geq \Pr\{\mathsf{reject} \mid \mathsf{H}_1\} = 1 - \Phi(-c)$. Under the alternative hypothesis H_1, each W_i has (in the worst case) distribution $\mathsf{Bernoulli}(1/2 + L_\infty(\mathbf{p},\mathbf{q})^2/2)$ by Theorem A.19, so that the empirical proportion

$\hat{B}$ is approximately normally distributed with mean at least $1/2 + L_\infty(\mathbf{p},\mathbf{q})^2/2$ and standard deviation at most $1/(2\sqrt{n})$. Under the alternative distribution of $\hat{B}$, the standard score c' of the critical value γ is

$$\begin{aligned} c' &= \frac{\gamma - (1/2 + L_\infty(\mathbf{p},\mathbf{q})^2/2)}{1/(2\sqrt{n})} \\ &= 2\sqrt{n}((1/2 + L_\infty(\mathbf{p},\mathbf{q})^2/4) - (1/2 + L_\infty(\mathbf{p},\mathbf{q})^2/2)) \\ &= -2\sqrt{n}(L_\infty(\mathbf{p},\mathbf{q})^2/4) \\ &= -\sqrt{n}L_\infty(\mathbf{p},\mathbf{q})^2/2 \\ &= -c, \end{aligned} \tag{36}$$

where the second equality follows from Eq. (35). Observe that the not reject region $F^c = [-\gamma,\gamma] \subset (-\infty,\gamma]$, and so the probability that $\hat{B}$ falls in F^c is at most the probability that $\hat{B} < \gamma$, which by Eq. (36) is equal to $\Phi(-c)$. It is then immediate that $\beta \geq 1 - \Phi(-c)$. □

The following corollary follows directly from Theorem 3.7.

Corollary A.21. *As the significance level α varies, the proposed test with ordering $\sqsubset$ and $m = 1$ achieves an overall error $(\alpha + (1-\beta))/2 \leq 3\Phi(-c)/2$ using $n = 4c^2/L_\infty(\mathbf{p},\mathbf{q})^4$ samples.*

A.5 Distribution of the test statistic under the alternative hypothesis

In this subsection we derive the distribution of R under the alternative hypothesis $\mathbf{p} \neq \mathbf{q}$. As before, write $\tilde{\mathbf{p}}(x) := \sum_{x' < x} \mathbf{p}(x)$.

Theorem A.22. *The distribution of R is given by*

$$\Pr\{R = r\} = \sum_{x \in \mathcal{T}} H(x,m,r)\,\mathbf{q}(x) \tag{37}$$

for $0 \leq r \leq m$, where $H(x,m,r) :=$

$$\begin{cases} \binom{r}{m} [\tilde{\mathbf{p}}(x)]^r [1-\tilde{\mathbf{p}}(x)]^{m-r} & (\mathbf{p}(x) = 0) \\ \dfrac{1}{m+1} & (\mathbf{p}(x) = 1) \\ \displaystyle\sum_{e=0}^{m} \Bigg\{ \Bigg[\sum_{j=0}^{e} \binom{m-e}{r-j} \left[\frac{\tilde{\mathbf{p}}(x)}{1-\mathbf{p}(x)}\right]^{r-j} \left[1 - \frac{\tilde{\mathbf{p}}(x)}{1-\mathbf{p}(x)}\right]^{(m-e)-(r-j)} \left(\frac{1}{e+1}\right) \Bigg] \binom{m}{e} [\mathbf{p}(x)]^m [1-\mathbf{p}(x)]^{e-m} \Bigg\} & (0 < \mathbf{p}(x) < 1) \end{cases}$$

Proof. Define the following random variables:

$$L := \sum_{i=1}^{m} \mathbb{I}[X_i \prec X_0], \tag{38}$$

$$E := \sum_{i=1}^{m} \mathbb{I}[X_i = X_0], \tag{39}$$

$$G := \sum_{i=1}^{m} \mathbb{I}[X_i \succ X_0]. \tag{40}$$

We refer to L, E, and G as "bins", where L is the "less than" bin, E is the "equal to" bin, and G is the "greater than" bin (all with respect to X_0). Total probability gives

$$\begin{aligned} \Pr\{R = r\} &= \sum_{x \in \mathcal{T}} \Pr\{R = r, X_0 = x\} \\ &= \sum_{\substack{x \in \mathcal{T} \\ \mathbf{q}(x) > 0}} \Pr\{R = r \mid X_0 = x\}\,\mathbf{q}(x). \end{aligned}$$

Fix $x \in \mathcal{T}$ such that $\mathbf{q}(x) > 0$. Consider $\Pr\{R = r \mid X_0 = s\}$. The counts in bins L, E, and G are binomial random variables with m trials, where the bin L has success probability $\tilde{\mathbf{p}}(x)$, the bin E has success probability $\mathbf{p}(x)$, and the bin G has success probability $1 - (\tilde{\mathbf{p}}(x) + \mathbf{p}(x))$. We now consider three cases.

Case 1: $\mathbf{p}(x) = 0$. The event $\{E = 0\}$ occurs with probability one since each X_i, for $1 \leq i \leq m$, cannot possibly be equal to x. Therefore, conditioned on $\{X_0 = x\}$, the event $\{R = r\}$ occurs if and only if $\{L = r\}$. Since L is binomially distributed,

$$\begin{aligned} \Pr\{R = r \mid X_0 = x\} &= \Pr\{L = r \mid X_0 = x\} \\ &= \binom{m}{r} [\tilde{\mathbf{p}}(x)]^r [1-\tilde{\mathbf{p}}(x)]^{m-r}. \end{aligned}$$

Case 2: $\mathbf{p}(x) = 1$. Then the event $\{E = m\}$ occurs with probability one since each X_i, for $1 \leq i \leq m$, can only equal s. The uniform numbers $U_0, \dots, U_m$ used to break the ties will determine the rank R of X_0. Let B be the rank of U_0 among the m other uniform random variables $U_1, \dots, U_m$. The event $\{R = r\}$ occurs if and only if $\{B = r\}$. Since the U_i are i.i.d., B is uniformly distributed over $\{0, 1, 2, \dots, m\}$ by Lemma A.2. Hence

$$\Pr\{R = r \mid X_0 = x\} = \Pr\{B = r \mid X_0 = x\} = \frac{1}{m+1}.$$

Case 3: $0 < \mathbf{p}(x) < 1$. By total probability,

$$\begin{aligned} &\Pr\{R = r \mid X_0 = x\} \\ &= \sum_{e=0}^{m} \Pr\{R = r \mid X_0 = x, E = e\} \Pr\{E = e \mid X_0 = x\}. \end{aligned}$$

Since E is binomially distributed,

$$\Pr\{E = e \mid X_0 = x\} = \binom{m}{e} [\mathbf{p}(x)]^e [1-\mathbf{p}(x)]^{m-e}.$$

We now tackle the event $\{R = r \mid X_0 = x, E = e\}$. The uniform numbers $U_0, \dots, U_m$ used to break the ties will determine the rank R of X_0. Define B to be the rank of U_0 among the e other uniform random variables assigned to bin E, i.e., those U_i for $1 \le i \le m$ such that $X_i = s$. The random variable B is independent of all the X_i, but is dependent on E. Given $\{E = e\}$, B is uniformly distributed on $\{0, 1, \dots, e\}$. By total probability,

$$\begin{aligned}
&\Pr\{R = r \mid X_0 = x, E = e\} \\
&\quad= \sum_{b=0}^{e} \big[\Pr\{R = r \mid X_0 = x, E = e, B = b\} \\
&\qquad\qquad \Pr\{B = b \mid E = e\}\big] \\
&\quad= \sum_{b=0}^{e} \Pr\{R = r \mid X_0 = x, E = e, B = b\} \frac{1}{e+1}.
\end{aligned}$$

Conditioned on $\{E = e\}$ and $\{B = 0\}$, the event $\{R = r\}$ occurs if and only if $\{L = r\}$, since exactly 0 random variables in bin E "are less" than X_0, so exactly r random variables in bin L are needed to ensure that the rank of X_0 is r. By the same reasoning, for $0 \le b \le e$, conditioned on $\{E = e, B = b\}$ we have $\{R = r\}$ if and only if $\{L = r - b\}$.

Now, conditioned on $\{E = e\}$, there are $m - e$ remaining assignments to be split among bins L and G. Let i be such that $X_i \ne x$. Then the relative probability that X_i is assigned to bin L is $\tilde{\mathbf{p}}(x)$ and to bin G is $1 - (\tilde{\mathbf{p}}(x) + \mathbf{p}(x))$. Renormalizing these probabilities, we conclude that L is conditionally (given $\{E = e\}$) a binomial random variable with $m - e$ trials and success probability $\tilde{\mathbf{p}}(x)/(\tilde{\mathbf{p}}(x) + (1 - (\tilde{\mathbf{p}}(x) + \mathbf{p}(x)))) = \tilde{\mathbf{p}}(x)/(1 - \mathbf{p}(x))$. Hence

$$\begin{aligned}
&\Pr\{R = r \mid X_0 = x, E = e, B = b\} \\
&= \Pr\{L = r - b \mid X_0 = x, E = e\} \\
&= \binom{m-e}{r-j} \left[\frac{\tilde{\mathbf{p}}(x)}{1-\mathbf{p}(x)}\right]^{r-j} \left[1-\frac{\tilde{\mathbf{p}}(x)}{1-\mathbf{p}(x)}\right]^{(m-e)-(r-j)},
\end{aligned}$$

completing the proof. □

Remark A.23. The sum in Eq. (37) of Theorem A.22 converges since $H(x, m, r) \le 1$.

Remark A.24. Theorem A.22 shows that it is not the case that we must have $\mathbf{p} = \mathbf{q}$ whenever there exists some m for which the rank R is uniform on $[m + 1]$. For example, let $m = 1$, let $\mathcal{T} \coloneqq \{0, 1, 2, 3\}$, let $\prec$ be the usual order $<$ on $\mathcal{T}$, and let $\mathbf{p} \coloneqq \frac{1}{2}\delta_0 + \frac{1}{2}\delta_3$ and $\mathbf{q} \coloneqq \frac{1}{2}\delta_1 + \frac{1}{2}\delta_2$. Let $X \sim \mathbf{p}$ and $Y \sim \mathbf{q}$. Then we have $\Pr\{R = 0\} = \Pr\{X > Y\} = 1/2 = \Pr\{Y < X\} = \Pr\{R = 1\}$. Rather, Theorem A.1 tells us merely if R is not uniform on $\{0, \dots, m\}$ for *some* m, then $\mathbf{p} \ne \mathbf{q}$. In the example given above, $m = 2$ (and so by Theorem A.7 all $m \ge 2$) provides such a witness.