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:

\[
\begin{align*}
X_0 & \sim \mathbf{q} \quad (13) \\
U_0 & \sim \text{Uniform}(0, 1) \quad (14) \\
X_1, X_2, \ldots, X_m & \sim_{\text{iid}} \mathbf{p} \quad (15) \\
U_1, U_2, \ldots, U_m & \sim_{\text{iid}} \text{Uniform}(0, 1) \quad (16) \\
R & = \sum_{j=1}^{m} I[X_j < X_0] + I[X_j = X_0, U_j < U_0]. \quad (17)
\end{align*}
\]

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 $[m+1] := \{0, 1, \ldots, m\}$.

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 $S_i := \sum_{j=0}^m I[T_j < T_i]$ 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 the set $\{0, 1/m, 2/m, \ldots, 1\}$, 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.,

\[
\lim_{m \to \infty} \Pr\{V_m < u\} = \Pr\{U < u\} = u. \quad (18)
\]

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

\[
F_m(u) = \begin{cases} 
1/(m+1) & u \in [0, 1/m) \\
2/(m+1) & u \in [1/m, 2/m) \\
\vdots \\
(m/(m+1) & u \in [(m-1)/m, 1/m) \\
1 & u \in [1/m, 1]
\end{cases}
\]

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 $p: (\mathbb{Q} \cap [0, 1]) \to [0, 1]$ 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 \(\{q_1, q_2, \ldots, q_m, \ldots\}\) 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 $\sum_{a < x \leq b} p(x)$ 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 \( T \) is countable, by a standard back-and-forth argument the total order \( (T, \prec) \) is isomorphic to \( (B, <) \) for some subset \( B \subseteq \mathbb{Q} \cap (0, 1) \). Without loss of generality, we may therefore take \( T \) to be \( \mathbb{Q} \cap [0, 1] \) and assume that \( p(0) = p(1) = 0 \).

Consider the unit square \([0, 1]^2\) equipped with the dictionary order \( \triangleq_d \). This is a total order with the least upper bound property. For each \( i \in \{m + 1\} \), define \( T_i := (X_i, U_i) \), 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 \( \sum_{i=0}^m \mathbb{I}[T_i \triangleq_d T_0] \) of \( T_0 \) taken according to the dictionary order.

(Necessity) Suppose \( p = q \). Then \( T_0, \ldots, T_m \) are independent and identically distributed. Since \( U_0, \ldots, U_m \) are continuous random variables, we have \( \mathbb{P}(T_i = T_j) = 0 \) 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 \([0, 1, 2, \ldots, m]\). We begin the proof by first constructing a distribution function \( F_p \) on the unit square and then establishing several of its properties. First let \( \hat{p} : [0, 1] \to [0, 1] \) be the “left-closed right-open” cumulative distribution function of \( p \), defined by

\[
\hat{p}(x) := \sum_{y \in \mathbb{Q} \cap (0, x]} p(y)
\]

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

\[
F_p(x, u) := \hat{p}(x) + up'(x)
\]

for \( (x, u) \in [0, 1]^2 \). To establish that \( F_p \) 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 \( F_p(0, 0) = 0 \) and \( F_p(1, 1) = 1 \). Furthermore, To establish that \( F_p \) is monotonically non-decreasing, put \( x < y \) and \( u < v \) and observe that

\[
F_p(x, u) = \hat{p}(x) + up'(x)
\]

\[
\leq \hat{p}(x) + p'(x)
\]

\[
\leq \sum_{z \in \mathbb{Q} \cap (0, y]} p'(z)
\]

\[
= \hat{p}(y)
\]

\[
\leq F_p(y, u)
\]

and

\[
F_p(x, u) = \hat{p}(x) + up'(x)
\]

\[
\leq \hat{p}(x) + pv'(x)
\]

\[
= F_p(x, v).
\]

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

\[
\epsilon > F(x', u) - F(x, u)
\]

\[
= \hat{p}(x') + up'(x') - \hat{p}(x) - up(x)
\]

\[
= \hat{p}(x') + p'(x') - \hat{p}(x) - p(x)
\]

\[
= \sum_{y \in \mathbb{Q} \cap (x, x'] } p(y),
\]

which is immediate from the Cauchy criterion.

Finally, we note that Lemma A.4 and the continuity of \( F_p \) in \( u \) together imply that \( F_p \) 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 \( F_p^{-1} : [0, 1] \to [0, 1]^2 \) by

\[
F_p^{-1}(s) := \inf \{ (x, u) \mid F_p(x, u) = s \} \quad (19)
\]

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

Now define the rank function

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

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

\[
\lim_{m \to \infty} \mathbb{P} \left\{ \frac{1}{m} r(T_0, \{T_1, \ldots, T_m\}) < s \right\}
\]

\[
= \mathbb{P} \{ U_0 < s \}
\]

\[
= s. \quad (20)
\]

for \( s \in [0, 1] \).

For any \( t \in [0, 1] \) and \( m \geq 1 \), the random variable \( \hat{F}_m(t) := \hat{r}(t, \{T_1, \ldots, T_m\})/m \) is the empirical distribution function of \( F_p \). 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 \( \{\hat{F}_m(t)\}_{m \geq 1} \) converges a.s. to the real number \( F_p(t) \) uniformly in \( t \). Hence the sequence \( \{F_m(T_0)\}_{m \geq 1} \) converges a.s. to the
random variable $\hat{F}_p(T_0)$, so that for any $s \in [0,1]$,

$$\lim_{m \to \infty} \Pr \left\{ \frac{1}{m} \hat{F}_p(T_0, \{T_1, \ldots, T_m\}) < s \right\} = \lim_{m \to \infty} \Pr \left\{ \hat{F}_p(T_0) < s \right\} = \Pr \{ F_p(T_0) < s \}$$

for $s \in [0,1]$. Since $0 \leq F_p(x,u) \leq 1$, for each $(x,u) \in [0,1]^2$ we have

$$F_q^{-1}(F_p(x,u)) = F_p^{-1}(F_q(x,u))$$

for $(x,u) \in [0,1]^2$. Fixing $u = 0$, we obtain

$$\hat{p}(x) = F_p(x,0) = F_q(x,0) = q(x)$$

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

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

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

Then we have

$$\hat{q}(b) = q(a) + q(x) + \sum_{x \in Q(a,b)} q(x) = \hat{p}(a) + q(a) + \sum_{x \in Q(a,b)} q(x) < \hat{p}(a) + q(a) + (p(a) - q(a)) = \hat{p}(a) + p(a) \leq \hat{p}(b),$$

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

The following corollary is an immediate consequence.

**Corollary A.6** (Corollary 3.3 in the main text). If $p \neq q$, 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 $p \neq q$, 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 $Z_1, \ldots, Z_{m+1}$ is a finitely exchangeable sequence of Bernoulli random variables. If $S_m := \sum_{i=1}^{m} Z_i$ is not uniformly distributed on $[m+1]$, then $S_{m+1} := \sum_{i=1}^{m+1} Z_i$ 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 Bernoulli($r$). There are two cases.

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

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

and so $S_{\ell}$ is not uniformly distributed on $[\ell + 1]$. In particular, this holds for $\ell$ 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

$$\Pr \{ S_m = k \} = \Pr \{ S_m = k \text{ and } Z_{m+1} = 0 \} + \Pr \{ S_m = k \text{ and } Z_{m+1} = 1 \}.$$  

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 $(m)_{k}$ assignments of $(Z_1, \ldots, Z_m)$ that have exactly $k$ ones and $Z_{m+1} = 0$. All such assignments are disjoint events. Define the event

$$A := \{ Z_1 = \cdots = Z_k = 1 \text{ and } Z_{k+1} = \cdots = Z_m = Z_{m+1} = 0 \}.$$
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 $\{S_m = k\}$ is the union of all $\binom{m+1}{k}$ assignments of $(Z_1, \ldots, Z_{m+1})$ that have exactly $k$ ones. All the assignments are disjoint events and each has probability $\Pr \{A\}$, so

$$\Pr \{S_m = k\} = \binom{m+1}{k} \Pr \{A\} = \frac{1}{m+2}. \quad (28)$$

Second, the event $\{S_m = k\} \cap \{Z_{m+1} = 1\}$ is the union over all $\binom{m+1}{k}$ assignments of $(Z_1, \ldots, Z_m)$ that have exactly $k$ ones and also $Z_{m+1} = 1$. All such assignments are disjoint events. Define the event $B := \{Z_1 = \cdots = Z_k = Z_{m+1} = 1 \text{ and } Z_{k+1} = \cdots = Z_m = 0\}$.

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 $(Z_1, \ldots, Z_{m+1})$ that have exactly $k+1$ ones. All the assignments are disjoint events and each has probability $\Pr \{B\}$, so

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

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

$$\frac{\Pr \{S_m = k\}}{1/(m+2)} = \frac{\Pr \{S_m = k \text{ and } Z_{m+1} = 0\}}{1/(m+2)} + \frac{\Pr \{S_m = k \text{ and } Z_{m+1} = 1\}}{1/(m+2)} = \binom{m}{k} \Pr \{A\} + \binom{m}{k+1} \Pr \{B\}$$

$$= \frac{m!}{k!(m-k)!} \frac{k!(m+1-k)!}{(m+1)!} \frac{m!(k+1)(m+1-(k+1))}{k!(m-k)!} \frac{1}{(m+2)}.$$

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

Now, each indicator $Z_i := I[T_i < T_0]$ is a Bernoulli random variable, and they are identically distributed since $(T_1, \ldots, T_{M+1})$ is an i.i.d. sequence. Furthermore the sequence $(Z_1, \ldots, Z_{M+1})$ is finitely exchangeable since the $Z_i$ are conditionally independent given $T_0$. Then the sequence of indicators $(I[T_i < T_0], I[T_2 < T_0], \ldots, I[T_{M+1} < T_0])$ 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 $p$ and $q$ satisfy an adversarial symmetry relationship under the selected ordering $\prec$, the rank is non-uniform for any choice of $m \geq 1$. Let $\prec$ denote the lexicographic order on $T \times [0,1]$ induced by $(T, \prec)$ and $([0,1], \prec)$.

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

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

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

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

Define

$$A := \{x \in T \mid q(x) > p(x)\}$$

to be the set of all elements of $T$ that have a greater probability according to $q$ than according to $p$, and let $A^c$ denote its complement. Let $h_{p,q}$ be the signed measure given by the difference $h_{p,q}(x) := q(x) - p(x)$
We first establish some easy lemmas.

With ties broken arbitrarily. The smallest element in

\[ \text{Lemma A.12.} \quad \tilde{q}(x) > \tilde{p}(x) \quad \text{for all } x \in \mathcal{T}. \]

Proof. Let \( \mathcal{T}_x := \{ y \in \mathcal{T} \mid y \prec x \} \). If \( x \in A \) then \( \mathcal{T}_x \subseteq A \), and so

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

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 \( A_x^\prec := \{ y \in A^\prec \mid y \prec x \} \). Then

\[ \tilde{q}(x) - \tilde{p}(x) = \sum_{y \prec x} [q(y) - p(y)] \]

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:

\[ X_p \sim p \]
\[ Y_q \sim q \]
\[ R_{p,q} | X_p, Y_q \sim \begin{cases} 
0 & \text{if } X_p \succ Y_q, \\
1 & \text{if } X_p \prec Y_q, \\
\text{Bernoulli}(1/2) & \text{if } X_p = Y_q.
\end{cases} \]

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

In other words, the procedure samples \( X_p \sim p \) and \( Y_q \sim q \) independently. Given these values, it then sets \( R_{p,q} \) to be 0 if \( X_p > Y_q \), to be 1 if \( X_p < Y_q \), 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 \( p \) and \( q \) (and accordingly will retain the subscripts).

We now prove the following theorem.

**Theorem A.13** (Theorem 3.6 in the main text). If \( p \neq q \), 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

\[ \Pr \{ R = 0 \} = \sum_{x,y \in \mathcal{T}} \Pr \{ R = 0 \mid X=x, Y=y \} \Pr \{ Y = y \} \Pr \{ X = x \} \]

\[ = \sum_{x,y \in \mathcal{T}} \Pr \{ R = 0 \mid X=x, Y=y \} q(y)p(x) \]

\[ = \sum_{x \in \mathcal{T}} \Pr \{ R = 0 \mid X=x, Y=x \} q(x)p(x) \]

\[ + \sum_{y < x \in \mathcal{T}} \Pr \{ R = 0 \mid X=x, Y=y \} q(y)p(x) \]

\[ = \sum_{y \in \mathcal{T}} [q(y) - p(y)] \]

\[ = \sum_{y \in A^\prec} [q(y) - p(y)] + \sum_{y \in A^\prec} [q(y) - p(y)] \]

\[ = \sum_{y \in A^\prec} [q(y) - p(y)] \]

\[ = \sum_{y \in A^\prec} [q(y) - p(y)] - \sum_{y \in A^\prec} [p(y) - q(y)] \]

\[ \geq 0, \]

plugging (i).

\[ \sum_{x \in \mathcal{T}} q(x) - \sum_{x \in \mathcal{T}} p(x) = 0, \]

as desired.

Given a probability distribution \( r \), define its cumulative distribution function \( \tilde{r} \) by \( \tilde{r}(x) := \sum_{y < x} r(y) \).

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

Proof. Let \( \mathcal{T}_x := \{ y \in \mathcal{T} \mid y \prec x \} \). If \( x \in A \) then \( \mathcal{T}_x \subseteq A \), and so

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

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 \( A_x^\prec := \{ y \in A^\prec \mid y \prec x \} \). Then

\[ \tilde{q}(x) - \tilde{p}(x) = \sum_{y \prec x} [q(y) - p(y)] \]
We have just exhibited an ordering such that when \( p \) and \( q \) satisfy:

\[
\sum_{x \in \mathcal{T}} \Pr \{ R = 0 | X = x, Y = y \} q(y)p(x)
\]

\[
= \frac{1}{2} \sum_{x \in \mathcal{T}} q(x)p(x) + 1 \sum_{y, x \in \mathcal{T}} q(y)p(x)
\]

\[
+ 0 \sum_{x \in \mathcal{T}} q(y)p(x)
\]

\[
= \frac{1}{2} \sum_{x \in \mathcal{T}} p(x)q(x) + \sum_{x \in \mathcal{T}} \tilde{p}(x)p(x).
\]

An identical argument establishes that:

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

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

\[
\Pr \{ R = 0 \} - \Pr \{ R = 1 \}
\]

\[
= \sum_{x \in \mathcal{T}} \tilde{q}(x)p(x) - \sum_{x \in \mathcal{T}} \tilde{p}(x)q(x)
\]

\[
> \sum_{x \in \mathcal{T}} \tilde{p}(x)p(x) - \sum_{x \in \mathcal{T}} \tilde{p}(x)q(x)
\]

\[
= \sum_{x \in \mathcal{T}} \tilde{p}(x)[p(x) - q(x)]
\]

\[
= \sum_{x \in \mathcal{A}} [\max_{y \in \mathcal{A}} \tilde{p}(y)](q(x) - p(x))
\]

\[
= \sum_{x \in \mathcal{A}} [\max_{y \in \mathcal{A}} \tilde{p}(y)](q(x) - p(x))
\]

\[
- \sum_{x \in \mathcal{A}} \tilde{p}(x)(q(x) - p(x))
\]

\[
= \sum_{x \in \mathcal{A}} [\max_{y \in \mathcal{A}} \tilde{p}(y)](q(x) - p(x))
\]

\[
- \sum_{x \in \mathcal{A}} \tilde{p}(x)(q(x) - p(x))
\]

\[
> 0.
\]

The first inequality follows from Lemma A.12; the second inequality follows from monotonicity of \( \tilde{p} \); 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. \( \square \)

### A.3 A tighter bound in terms of \( L_\infty(p, q) \)

We have just exhibited an ordering such that when \( p \neq q \) 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 \( L_\infty \) distance between \( p \) and \( q \).

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 \( p, q \) and all \( \delta > 0 \) there is an \( \epsilon > 0 \) such that for all distributions \( p' \) on \( \mathcal{T} \) with \( \sup_{x \in \mathcal{T}} |p(x) - p'(x)| < \epsilon \), we have:

\[
|\Pr(R_{p,q} = 0 | X_p \in B, Y_q \in C) - \Pr(R_{p',q} = 0 | X_{p'} \in B, Y_q \in C)| < \delta.
\]

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

\[
|h_{p,q}(a) - h_{p,q}(b)| \geq \epsilon.
\]

From Lemma A.14 we immediately obtain the following.

**Lemma A.16.** For all \( p, q \) and all \( \delta > 0 \) there is an \( \epsilon > 0 \) and an \( \epsilon \)-discrete distribution \( p' \) on \( \mathcal{T} \) such that for all \( B, C \subseteq \mathcal{T} \),

\[
|\Pr(R_{p,q} = 0 | X_p \in B, Y_q \in C) - \Pr(R_{p',q} = 0 | X_{p'} \in B, Y_q \in C)| < \delta.
\]

The next lemma will be crucial for proving our bound.

**Lemma A.17.** Let \( p_0 \) and \( p_1 \) be probability measures on \( \mathcal{T} \), and let \( \prec \) be a total order on \( \mathcal{T} \) such that if \( h_{p_0,q}(x) > h_{p_0,q}(x') \) then \( x \prec x' \) and if \( h_{p_0,q}(x) < h_{p_0,q}(x') \) then \( x \succ x' \). Suppose that if \( h_{p_0,p_1}(x) > 0 \) and \( h_{p_0,p_1}(y) \leq 0 \), then \( x \prec y \). Then \( \Pr(R_{p_0,q} = 0) \geq \Pr(R_{p_1,q} = 0) \).

**Proof.** Note that:

\[
\Pr(R_{p_1,q} = 0 | Y_q = y)
\]

\[
= \sum_{x \succ y} p_1(x) + \frac{1}{2} p_1(y)
\]

\[
= \sum_{x \succ y} p_0(x) + h_{p_0,p_1}(x) + \frac{1}{2} [p_0(y) + h_{p_0,p_1}(y)]
\]

\[
= \Pr(R_{p_0,q} = 0 | Y_q = y) + \sum_{x \succ y} h_{p_0,p_1}(x) + \frac{1}{2} h_{p_0,p_1}(y)
\]

\[
= \Pr(R_{p_0,q} = 0 | Y_q = y) - \sum_{x \prec y} h_{p_0,p_1}(x) - \frac{1}{2} h_{p_0,p_1}(y),
\]

where the last equality holds because \( \sum_{x \in \mathcal{T}} h_{p_0,p_1}(x) = 0 \). But by our assumption, we know that \( \sum_{x \prec y} h_{p_0,p_1}(x) + \frac{1}{2} h_{p_0,p_1}(y) \) is non-negative and so \( \Pr(R_{p_1,q} = 0 | Y_q = y) \leq \Pr(R_{p_0,q} = 0 | Y_q = y) \), from which the result follows. \( \square \)

We will now provide a lower bound on \( \Pr(R_{p,q} = 0) \).
Proposition A.18.

\[
\Pr(R_{p,q} = 0) \geq \frac{1}{2} + \frac{1}{2} \max_{x \in \mathcal{T}} h_{p,q}(x)^2 .
\]  

(31)

Proof. Recall that \( A := \{ x \in \mathcal{T} \mid q(x) > p(x) \} \). 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 \( k := |A| \). Further, by Lemma A.16 we may assume without loss of generality that \( p, q \) are an \( \epsilon \)-discrete pair (for some fixed but small \( \epsilon \)) with \( |\mathcal{T} \setminus \epsilon| < L_\infty(p, q) \). Let \( (x^+_0, \ldots, x^+_k) \) be the collection \( A \) listed in \( \epsilon \)-increasing order. Let \( (x^-_0, \ldots, x^-_{k-1}) \) be the collection \( \mathcal{T} \setminus A \) listed in \( \epsilon \)-increasing order.

Let \( p^* \) be any probability measure such that

\[
p^*(x) = \begin{cases} p(x) - \epsilon \ell & (x = x^-_\ell; \epsilon \ell \geq 0), \\ q(x) - (k - \ell) \cdot \epsilon & (x = x^+_\ell; 0 \leq \ell < k - 1), \\ p(x) & (x = x^+_k). \end{cases}
\]

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

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

(a) \( x < y \) if and only if \( h_{p^*, q}(x) > h_{p^*, q}(y) \),

(b) \( h_{p,q}(x^+_0) = h_{p^*, q}(x^+_0) \),

(c) if \( h_{p^*, q}(x) > 0 \) and \( h_{p^*, q}(y) \leq 0 \) then \( x < y \), and

(d) \( (p, q^*) \) is an \( \epsilon \)-discrete pair.

Note that \( \Pr(R_{p,q} = 0) \geq \Pr(R_{p^*, q} = 0) \), by Lemma A.17 and (c). For simplicity, let \( A_0 := \{ x^+_0 \} \), \( A_1 := \{ x^+_1 \}_{1 \leq i < k - 1} \) and \( D := \mathcal{T} \setminus A \).

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

Case 1: \( Y_q = x^-_i \). We have

\[
\Pr(R_{p^*, q} = 0 \mid Y_q = x^-_i) = \sum_{i < \ell < k} p^*(x^-_\ell) + \frac{1}{2} p^*(x^-_i).
\]

Case 2: \( Y_q \in A_1 \). We have

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

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

Case 3: \( Y_q \in A_0 \). We have

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

We may calculate these terms as follows:

\[
\begin{align*}
p^*(D) &= q(D) + h_{p,q}(x^+_0) + (k(k-1)/2)\epsilon, \\
p^*(A_1) &= q(A_1) - (k(k-1)/2)\epsilon, \\
p^*(A_0) &= q(A_0) - h_{p,q}(x^+_0).
\end{align*}
\]

Putting all of this together, we obtain

\[
\Pr(R_{p^*, q} = 0)
\]

\[
= \sum_{i < k} \sum_{i < \ell < k} q(x^-_i) p^*(x^-_\ell) + \frac{1}{2} \sum_{i < k} q(x^-_i) p^*(x^-_i)
\]

\[
+ q(A_1) p^*(D) + \frac{1}{2} q(A_1) p^*(A_1) + q(A_1) f_0(\epsilon)
\]

\[
+ q(A_0) p^*(A_1) + q(A_0) p^*(D) + \frac{1}{2} q(A_0) p^*(A_0)
\]

\[
= \sum_{i < k} \sum_{i < \ell < k} q(x^-_i) [q(x^-_\ell) - h_{p^*, q}(x^-_\ell)]
\]

\[
+ \frac{1}{2} \sum_{i < k} q(x^-_i) [q(x^-_\ell) - h_{p^*, q}(x^-_\ell)]
\]

\[
+ q(A_1) [q(D) + h_{p,q}(x^+_0)] + q(A_1) h_{p,q}(x^+_0)
\]

\[
+ q(A_0) [q(A_1) + q(A_0) [q(D) + h_{p,q}(x^+_0)]
\]

\[
+ \frac{1}{2} q(A_0) [q(A_0) - h_{p,q}(x^+_0)] + f_1(\epsilon)
\]

\[
= \sum_{i < k} \sum_{i < \ell < k} q(x^-_i) [q(x^-_\ell) + \frac{1}{2} q(x^-_i) q(x^-_\ell)]
\]

\[
+ q(A_1) [q(D) + \frac{1}{2} q(A_1) q(A_1) + q(A_0) q(A_1)]
\]

\[
+ q(A_0) [q(D) + \frac{1}{2} q(A_0) q(A_0)]
\]

\[
- \sum_{i < k} \sum_{i < \ell < k} q(x^-_i) h_{p^*, q}(x^-_\ell)
\]

\[
- \frac{1}{2} \sum_{i < k} q(x^-_i) h_{p^*, q}(x^-_\ell) + q(A_1) h_{p,q}(x^+_0)
\]

\[
+ q(A_0) h_{p,q}(x^+_0) - \frac{1}{2} q(A_0) h_{p,q}(x^+_0) + f_1(\epsilon)
\]

\[
= \sum_{i < k} \sum_{i < \ell < k} q(x^-_i) [q(x^-_\ell) + \frac{1}{2} q(x^-_i) q(x^-_\ell)]
\]

\[
+ q(A_1) [q(D) + \frac{1}{2} q(A_1) q(A_1) + q(A_0) q(A_1)]
\]

\[
+ q(A_0) [q(D) + \frac{1}{2} q(A_0) q(A_0)]
\]

\[
- \sum_{i < k} \sum_{i < \ell < k} q(x^-_i) h_{p^*, q}(x^-_\ell)
\]

\[
- \frac{1}{2} \sum_{i < k} q(x^-_i) h_{p^*, q}(x^-_\ell) + q(A_1) h_{p,q}(x^+_0)
\]

\[
+ q(A_0) h_{p,q}(x^+_0) - \frac{1}{2} q(A_0) h_{p,q}(x^+_0) + f_1(\epsilon),
\]
where \( f_1 \) is a function satisfying \( \lim_{\epsilon \to 0} f_1(\epsilon) = 0 \).

We also have
\[
\frac{1}{2} = \Pr(R_{q,q} = 0) = \sum_{i<k} \sum_{\ell<k} q(x_i^-)q(x_\ell^-) + \frac{1}{2} \sum_{i<k} q(x_i^-)q(x_i^+) \\
+ q(A_1)q(D) + \frac{1}{2} q(A_1)q(A_1) \\
+ q(A_0)q(A_1) + q(A_0)q(D) + \frac{1}{2} q(A_0)q(A_0).
\]

Putting these two equations together, we obtain
\[
\Pr(R_{p^*,q} = 0) = \frac{1}{2} - \Pr(R_{p,q} = 0)
= - \sum_{i<k} \sum_{\ell<k} q(x_i^-)h_{p^*,q}(x_\ell^-) \\
- \frac{1}{2} \sum_{i<k} q(x_i^-)h_{p,q}(x_i^+) + q(A_1)h_{p,q}(x_0^+) \\
+ \frac{1}{2} q(A_0)h_{p,q}(x_0^+) + f_1(\epsilon)
\geq \frac{1}{2} q(A_0)h_{p,q}(x_0^+) + f_1(\epsilon),
\]
as \( h_{p,q}(x^-) \leq 0 \) for all \( \ell < k \) and \( h_{p,q}(x_0^+) \leq 0 \).

But we know that
\[
q(A_0) = q(x_0^+) = p^*(x_0^+) + h_{p,q}(x_0^+) \geq h_{p,q}(x_0^+).
\]

Therefore, \( h_{p,q}(x_0^+) \) is the maximal value of \( h_{p,q} \), by taking the limit as \( \epsilon \to 0 \) we obtain
\[
\Pr(R_{p^*,q} = 0) = \frac{1}{2} + \frac{1}{2} \max_{x \in T} h_{p,q}(x)^2,
\]
as desired. \( \square \)

Finally, we arrive at the following theorem.

**Theorem A.19.** Given probability measure \( p,q \) on \( T \) there is a linear ordering \( \sqsubset \) of \( T \) such that if \( X_p \) and \( Y_q \) are sampled independently from \( p \) and \( q \) respectively then
\[
\Pr(X_q \sqsubset Y_p) \geq \frac{1}{2} + \frac{1}{2} L_\infty(p,q)^2.
\]

**Proof.** Note that
\[
L_\infty(p,q) = \max_{x \in T} \{ \max_{x \in T} h_{p,q}(x), \max_{x \in T} h_{q,p}(x) \}.
\]

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

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

**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 \( p = q \).

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
\[
n \approx 4c^2 / L_\infty(p,q)^4
\]
samples from \( q \), where \( \Phi \) 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 \( L_\infty = \max_{x \in T}(q(x) - p(x)) \). Let \( (Y_1, \ldots, Y_n) \sim_{iid} q \) be the \( n \) samples from \( q \). With \( m = 1 \), the testing procedure generates \( n \) samples \( (X_1, \ldots, X_n) \sim_{iid} p \), and \( 2n \) uniform random variables \( (U_1^1, \ldots, U_n^1, U_1^2, \ldots, U_n^2) \sim_{iid} \text{Uniform}(0,1) \) to break ties. Let \( \triangleq \) denote the lexicographic order on \( T \times [0,1] \) induced by \( (T, \triangleq) \) and \( (0,1], \triangleq) \). Define \( W_i := I[[Y_i, U_i^1] \triangleq (X_i, U_i^2)] \) for \( 1 \leq i \leq n \), to be the rank of the \( i \)-th observation from \( q \).

Under the null hypothesis \( H_0 \), each \( W_i \) has distribution \( \text{Bernoulli}(1/2) \) 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 \( \{W_1, \ldots, W_n\} \). Let \( \bar{B} := \sum_{i=1}^n (1 - W_i)/n \) denote the empirical proportion of zeros. By the central limit theorem, for sufficiently large \( n \), we have that \( \bar{B} \) 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 \( F = \{\gamma, \gamma\} \cup \{\gamma, \gamma\} \), where the critical value \( \gamma \) satisfies
\[
c = \frac{\gamma - 1/2}{1/(2\sqrt{n})} = 2\sqrt{n}(\gamma - 1/2).
\]
Replacing \( n \) in Eq. (7), we obtain
\[
\gamma = 1/2 + c/(2\sqrt{n}) = 1/2 + c/(2(2c/L_\infty(p,q)^2))
\]
\[
= 1/2 + L_\infty(p,q)^2/4.
\]
This construction ensures that \( \Pr\{\text{reject} \mid H_0\} = \alpha \).

We now show that the test with this rejection region has power \( \beta \geq \Pr\{\text{reject} \mid H_1\} = 1 - \Phi(-c) \). Under the alternative hypothesis \( H_1 \), each \( W_i \) has (in the worst case) distribution \( \text{Bernoulli}(1/2 + L_\infty(p,q)^2/2) \) by Theorem A.19, so that the empirical proportion
\( \hat{B} \) is approximately normally distributed with mean at least \( 1/2 + L_\infty(p, q)^2/2 \) and standard deviation at most \( 1/(2\sqrt{n}) \). Under the alternative distribution of \( \hat{B} \), the standard score \( c' \) of the critical value \( \gamma \) is

\[
c' = \frac{\gamma - (1/2 + L_\infty(p, q)^2/2)}{1/(2\sqrt{n})} = 2\sqrt{n}(1/2 + L_\infty(p, q)^2/2) - (1/2 + L_\infty(p, q)^2/2)) = -2\sqrt{n}L_\infty(p, q)^2/4 = -\sqrt{n}L_\infty(p, q)^2/2 = -c, \tag{36}
\]

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 \( \alpha \) varies, the proposed test with ordering \( \square \) and \( m = 1 \) achieves an overall error \( (\alpha + (1 - \beta)) \) using \( n = 4c^2L_\infty(p, q)^4 \) 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 \( p \neq q \). As before, write \( \tilde{p}(x) := \sum' \leq x p(x) \).

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

\[
\Pr \{ R = r \} = \sum_{x \in T} H(x, m, r) q(x) \tag{37}
\]

for \( 0 \leq r \leq m \), where \( H(x, m, r) := \begin{cases} \binom{r}{m} [\tilde{p}(x)]^r (1 - \tilde{p}(x))^{m-r} & (p(x) = 0) \\ \frac{1 - \tilde{p}(x)}{m + 1} & (p(x) = 1) \\ \binom{m}{j} \left[ \frac{1 - \tilde{p}(x)}{1 - \tilde{p}(x)} \right]^{r-j} \left[ 1 - \tilde{p}(x) \right]^{(m-j)-(r-j)} \left( \frac{1}{e + 1} \right)^j \left( \frac{m}{e} \right) [\tilde{p}(x)]^m [1 - \tilde{p}(x)]^{e-m} & (0 < p(x) < 1) \end{cases} \]

**Proof.** Define the following random variables:

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

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

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

\[
\Pr \{ R = r \} = \sum_{x \in T} \Pr \{ R = r, X_0 = x \} = \sum_{x \in T} \Pr \{ R = r \mid X_0 = x \} q(x).
\]

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

**Case 1:** \( p(x) = 0 \). 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,

\[
\Pr \{ R = r \mid X_0 = x \} = \Pr \{ L = r \mid X_0 = x \} = \binom{m}{r} [\tilde{p}(x)]^r [1 - \tilde{p}(x)]^{m-r}.
\]

**Case 2:** \( p(x) = 1 \). 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 \( U_0, \ldots, U_m \) 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 \( U_1, \ldots, U_m \). The event \( \{ R = r \} \) occurs if and only if \( \{ B = r \} \). Since the \( U_i \) are i.i.d., \( B \) is uniformly distributed over \( \{0, 1, 2, \ldots, m\} \) by Lemma A.2.

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

**Case 3:** \( 0 < p(x) < 1 \). By total probability,

\[
\Pr \{ R = r \mid X_0 = x \} = \sum_{e=0}^m \Pr \{ R = r \mid X_0 = x, E = e \} \Pr \{ E = e \mid X_0 = x \}.
\]

Since \( E \) is binomially distributed,

\[
\Pr \{ E = e \mid X_0 = x \} = \binom{m}{e} [p(x)]^e [1 - p(x)]^{m-e}.
\]
We now tackle the event \( \{ R = r \mid X_0 = x, E = e \} \). The uniform numbers \( U_0, \ldots, U_m \) 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 \( 1 \leq i \leq m \) 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 \( \{ 0, 1, \ldots, e \} \). By total probability,

\[
\Pr \{ R = r \mid X_0 = x, E = e \} = \sum_{b=0}^{e} \Pr \{ R = r \mid X_0 = x, E = e, B = b \} \Pr \{ B = b \mid E = e \} = \sum_{b=0}^{e} \Pr \{ R = r \mid X_0 = x, E = e, B = b \} \frac{1}{e+1}.
\]

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 \( 0 \leq b \leq e \), 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 \( X_i \neq x \). Then the relative probability that \( X_i \) is assigned to bin \( L \) is \( \hat{p}(x) \) and to bin \( G \) is \( 1 - (\hat{p}(x) + p(x)) \). Renormalizing these probabilities, we conclude that \( L \) is conditionally (given \( \{ E = e \} \)) a binomial random variable with \( m - e \) trials and success probability \( \hat{p}(x)/(\hat{p}(x) + (1 - (\hat{p}(x) + p(x)))) = \hat{p}(x)/p(x) \). Hence

\[
\Pr \{ R = r \mid X_0 = x, E = e, B = b \} = \Pr \{ L = r - b \mid X_0 = x, E = e \} = \binom{m - e}{r - j} \left( \frac{\hat{p}(x)}{1 - p(x)} \right)^{r - j} \left( \frac{1 - \hat{p}(x)}{1 - p(x)} \right)^{(m - e) - (r - j)},
\]

completing the proof.

\begin{remark}
The sum in Eq. (37) of Theorem A.22 converges since \( H(x, m, r) \leq 1 \).
\end{remark}

\begin{remark}
Theorem A.22 shows that it is not the case that we must have \( p = q \) whenever there exists some \( m \) for which the rank \( R \) is uniform on \( [m + 1] \).

For example, let \( m = 1 \), let \( T := \{ 0, 1, 2, 3 \} \), let \( \prec \) be the usual order \( \prec \) on \( T \), and let \( p := \frac{1}{2} \delta_0 + \frac{1}{2} \delta_3 \) and \( q := \frac{1}{4} \delta_1 + \frac{1}{4} \delta_2 \). Let \( X \sim p \) and \( Y \sim q \). 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 \( \{ 0, \ldots, m \} \) for some \( m \), then \( p \neq q \). In the example given above, \( m = 2 \) (and so by Theorem A.7 all \( m \geq 2 \)) provides such a witness.
\end{remark}