A NEW COEFFICIENT OF CORRELATION

SOURAV CHATTERJEE

Supplementary material: Proofs

A. Proof of Theorem 1.1

Throughout this proof and the rest of the manuscript, we will abbreviate $\xi_n(X,Y)$ as ξ_n and $\xi(X,Y)$ as ξ. For $t \in \mathbb{R}$, let $F(t) := \mathbb{P}(Y \le t)$ and $G(t) := \mathbb{P}(Y \ge t)$. Let μ be the law of Y. By the existence of regular conditional probabilities on regular Borel spaces (see for example [2, Theorem 2.1.15 and Exercise 5.1.16]), for each Borel set $A \subseteq \mathbb{R}$ there is a measurable map $x \mapsto \mu_x(A)$ from $\mathbb{R}$ into $[0,1]$, such that

(1) for any A, $\mu_X(A)$ is a version of $\mathbb{P}(Y \in A|X)$, and
(2) with probability one, μ_X is a probability measure on $\mathbb{R}$.

In the above sense, μ_x is the conditional law of Y given $X = x$. For each t, let $G_x(t) := \mu_x([t,\infty))$, and define

$$Q := \int \mathrm{Var}(G_X(t))d\mu(t). \tag{A.1}$$

(Since $t \mapsto \mathbb{E}(G_X(t))$ and $t \mapsto \mathbb{E}(G_X(t)^2)$ are both non-increasing maps, they are measurable. Therefore $t \mapsto \mathrm{Var}(G_X(t))$ is also measurable, and so the above integral is well-defined.)

Lemma A.1. *Let Q be as above. Then $Q = 0$ if and only if X and Y are independent.*

Proof. If X and Y are independent, then for any t, $\mathbb{P}(Y \ge t|X) = \mathbb{P}(Y \ge t)$ almost surely. Thus, $G_X(t) = G(t)$ almost surely, and so $\mathrm{Var}(G_X(t)) = 0$. Consequently, $Q = 0$.

Conversely, suppose that $Q = 0$. Then there is a Borel set $A \subseteq \mathbb{R}$ such that $\mu(A) = 1$ and $\mathrm{Var}(G_X(t)) = 0$ for every $t \in A$. Since $\mathbb{E}(G_X(t)) = G(t)$, $G_X(t) = G(t)$ almost surely for each $t \in A$. We claim that A can be chosen to be the whole of $\mathbb{R}$.

To show this, take any $t \in \mathbb{R}$. If $\mu(\{t\}) > 0$, then clearly t must be a member of A and there is nothing more to prove. So assume that $\mu(\{t\}) = 0$. This implies that G is right-continuous at t.

There are two possibilities. First, suppose that $G(s) < G(t)$ for all $s > t$. Then for each $s > t$, $\mu([t,s)) > 0$, and hence A must intersect $[t,s)$. This shows that there is a sequence r_n in A such that r_n decreases to t. Since

$G_X(r_n) = G(r_n)$ almost surely for each n, this implies that with probability one,

$$G_X(t) \ge \lim_{n\to\infty} G_X(r_n) = \lim_{n\to\infty} G(r_n) = G(t).$$

But $\mathbb{E}(G_X(t)) = G(t)$. Thus, $G_X(t) = G(t)$ almost surely.

The second possibility is that there is some $s > t$ such that $G(s) = G(t)$. Take the largest such s, which exists because G is left-continuous. If $s = \infty$, then $G(t) = G(s) = 0$, and hence $G_X(t) = 0$ almost surely because $\mathbb{E}(G_X(t)) = G(t)$. Suppose that $s < \infty$. Then either $\mu(\{s\}) > 0$, which implies that $G_X(s) = G(s)$ almost surely, or $\mu(\{s\}) = 0$ and $G(r) < G(s)$ for all $r > s$, which again implies that $G_X(s) = G(s)$ almost surely, by the previous paragraph. Therefore in either case, with probability one,

$$G_X(t) \ge G_X(s) = G(s) = G(t).$$

Since $\mathbb{E}(G_X(t)) = G(t)$, this implies that $G_X(t) = G(t)$ almost surely.

This completes the proof of our claim that for each $t \in \mathbb{R}$, $G_X(t) = G(t)$ almost surely. Therefore, for any $t \in \mathbb{R}$ and any Borel set $B \subseteq \mathbb{R}$,

$$\begin{aligned}\mathbb{P}(\{Y \ge t\} \cap \{X \in B\}) &= \mathbb{E}(\mathbb{P}(Y \ge t|X)1_{\{X\in B\}}) \\ &= G(t)\mathbb{P}(X \in B) = \mathbb{P}(Y \ge t)\mathbb{P}(X \in B).\end{aligned}$$

This proves that Y and X are independent. □

Corollary A.2. *If Y is not a constant, then $\int G(t)(1 - G(t))d\mu(t) > 0$.*

Proof. In Lemma A.1, take $X = Y$. Then $G_X(t) = 1_{\{X\ge t\}}$, and hence $\mathrm{Var}(G_X(t)) = G(t)(1 - G(t))$. But if Y is not a constant, then Y is not independent of itself. Hence Lemma A.1 implies that $Q > 0$, which gives what we want. □

Let $X_1, X_2, \ldots$ be an infinite sequence of i.i.d. copies of X. For each $n \ge 2$ and each $1 \le i \le n$, let $X_{n,i}$ be the element of the set $\{X_j : 1 \le j \le n, j \ne i\}$ that is immediately to the right of X_i. If there is no such element, then let $X_{n,i} = X_i$.

Lemma A.3. *With probability one, $X_{n,1} \to X_1$ as $n \to \infty$.*

Proof. Let ν be the law of X. Let A be the set of all $x \in \mathbb{R}$ such that $\nu([x, y)) > 0$ for any $y > x$. First, we show that $\nu(A^c) = 0$. Let K be the support of ν and let $B := A^c \cap K$. Since $\nu(K^c) = 0$, it suffices to show that $\nu(B) = 0$.

Take any $x \in B$. Since $x \in A^c$, there is some $y > x$ such that $\nu([x, y)) = 0$. For each $x \in B$, choose such a point y_x. We claim that the intervals $[x, y_x)$, as x ranges over B, are disjoint. To see this, take any distinct $x, x' \in B$, $x < x'$. If $[x, y_x)$ and $[x', y_{x'})$ are not disjoint, then $x' \in (x, y_x)$. But $\nu((x, y_x)) \le \nu([x, y_x)) = 0$. This contradicts the fact that $x' \in K$. Thus, we have established that the intervals $[x, y_x)$ are disjoint. But this implies that there can be at most countably many such intervals. Thus, B is at most countable. But for any $x \in B$, $\nu(\{x\}) \le \nu([x, y_x)) = 0$. This proves that $\nu(B) = 0$, and hence $\nu(A^c) = 0$.

Take any $\varepsilon > 0$. Let I be the interval $[X_1, X_1 + \varepsilon)$. Then

$$\mathbb{P}(|X_1 - X_{n,1}| \ge \varepsilon | X_1) \le (1 - \nu(I))^{n-1}.$$

Since $X_1 \in A$ almost surely, it follows that $\nu(I) > 0$ almost surely. Thus,

$$\lim_{n\to\infty} \mathbb{P}(|X_1 - X_{n,1}| \ge \varepsilon | X_1) = 0$$

almost surely, and hence

$$\lim_{n\to\infty} \mathbb{P}(|X_1 - X_{n,1}| \ge \varepsilon) = 0.$$

This proves that $|X_1 - X_{n,1}| \to 0$ in probability. But $|X_1 - X_{n,1}|$ is decreasing in n after the first time some X_j is drawn that is $\ge X_1$ (and there will always be such a time, since $\nu(I) > 0$). Therefore $|X_1 - X_{n,1}| \to 0$ almost surely. □

Lemma A.4. *For any measurable function* $f : \mathbb{R} \to [0, \infty)$,

$$\mathbb{E}(f(X_{n,1})) \le 2\mathbb{E}(f(X_1)).$$

Proof. Consider a particular realization of $X_1, \ldots, X_n$. In this realization, take any i and j such that $X_{n,i} = X_j$ and $X_j \ne X_i$. We claim that for any j, there can be at most one such i. Take any $k \notin \{i, j\}$. Then X_k cannot lie in the interval $[X_i, X_j)$, because that would contradict the fact that $X_{n,i} = X_j$. If $X_k < X_i$, then $X_{n,k} \ne X_j$ because X_i is closer to X_k on the right than X_j. On the other hand, if $X_k > X_j$, then obviously $X_{n,k} \ne X_j$. Thus, we conclude that for any j, there can be at most one i such that $X_{n,i} = X_j$ and $X_i \ne X_j$.

Now observe that since f is nonnegative,

$$\begin{aligned}\mathbb{E}(f(X_{n,i})) &\le \mathbb{E}(f(X_i)) + \mathbb{E}(f(X_{n,i})1_{\{X_{n,i} \ne X_i\}}) \\ &\le \mathbb{E}(f(X_i)) + \sum_{j=1}^n \mathbb{E}(f(X_j)1_{\{X_j = X_{n,i},\, X_j \ne X_i\}}).\end{aligned}$$

Combining the two observations and using symmetry, we get

$$\begin{aligned}\mathbb{E}(f(X_{n,1})) &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}(f(X_{n,i})) \\ &\le \frac{1}{n}\sum_{i=1}^n \mathbb{E}(f(X_i)) + \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}(f(X_j)1_{\{X_j = X_{n,i},\, X_j \ne X_i\}}) \\ &= \mathbb{E}(f(X_1)) + \frac{1}{n}\sum_{j=1}^n \mathbb{E}\biggl(f(X_j)\sum_{i=1}^n 1_{\{X_j = X_{n,i},\, X_j \ne X_i\}}\biggr) \\ &\le \mathbb{E}(f(X_1)) + \frac{1}{n}\sum_{j=1}^n \mathbb{E}(f(X_j)) = 2\mathbb{E}(f(X_1)),\end{aligned}$$

which completes the proof of the lemma. □

For the next result, we will need the following version of Lusin's theorem (proved, for example, by combining [4, Theorem 2.18 and Theorem 2.24]).

Lemma A.5 (Special case of Lusin's theorem)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a measurable function and ν be a probability measure on $\mathbb{R}$. Then, given any $\varepsilon > 0$, there is a compactly supported continuous function $g : \mathbb{R} \to \mathbb{R}$ such that $\nu(\{x : f(x) \neq g(x)\}) < \varepsilon$.*

Lemma A.6. *For any measurable $f : \mathbb{R} \to \mathbb{R}$, $f(X_1) - f(X_{n,1})$ tends to 0 in probability as $n \to \infty$.*

Proof. Fix some $\varepsilon > 0$. Let g be a function as in Lemma A.5, for the given f and ε, and $\nu =$ the law of X_1. Then note that for any $\delta > 0$,

$$\begin{aligned}&\mathbb{P}(|f(X_1) - f(X_{n,1})| > \delta)\\ &\le \mathbb{P}(|g(X_1) - g(X_{n,1})| > \delta) + \mathbb{P}(f(X_1) \neq g(X_1))\\ &\quad + \mathbb{P}(f(X_{n,1}) \neq g(X_{n,1})).\end{aligned}$$

By Lemma A.3 and the continuity of g,

$$\lim_{n\to\infty} \mathbb{P}(|g(X_1) - g(X_{n,1})| > \delta) = 0.$$

By the construction of g,

$$\mathbb{P}(f(X_1) \neq g(X_1)) < \varepsilon.$$

Finally, by Lemma A.4,

$$\mathbb{P}(f(X_{n,1}) \neq g(X_{n,1})) \le 2\mathbb{P}(f(X_1) \neq g(X_1)) \le 2\varepsilon.$$

Putting it all together, we get

$$\limsup_{n\to\infty} \mathbb{P}(|f(X_1) - f(X_{n,1})| > \delta) \le 3\varepsilon.$$

Since ε and δ are arbitrary, this completes the proof of the lemma. □

Let $\pi(i)$ be the rank of X_i, breaking ties at random so that π is a permutation of $\{1, \dots, n\}$. Define

$$N(i) := \begin{cases} \pi^{-1}(\pi(i) + 1) & \text{if } \pi(i) < n, \\ i & \text{if } \pi(i) = n. \end{cases}$$

We will now show that $\mathbb{P}(X_{n,1} = X_{N(1)}) \to 1$ as $n \to \infty$. For that, we need to recall the following formula.

Lemma A.7. *If $Z \sim \mathrm{Binomial}(m, p)$, then*

$$\mathbb{E}\biggl(\frac{1}{Z+1}\biggr) = \frac{1 - (1-p)^{m+1}}{(m+1)p}.$$

Proof. Let $x := p/(1-p)$. Then

$$\mathbb{E}\biggl(\frac{1}{Z+1}\biggr) = \sum_{k=0}^m \frac{1}{k+1}\binom{m}{k} p^k (1-p)^{m-k}$$

$$= \frac{(1-p)^m}{x} \sum_{k=0}^{m} \binom{m}{k} \frac{x^{k+1}}{k+1}$$
$$= \frac{(1-p)^m}{x} \int_0^x \sum_{k=0}^{m} \binom{m}{k} y^k dy$$
$$= \frac{(1-p)^m}{x} \int_0^x (1+y)^m dy = \frac{(1-p)^m}{x} \frac{(1+x)^{m+1}-1}{m+1}.$$

The result is obtained by substituting the value of x. □

Lemma A.8. *$\mathbb{P}(X_{n,1} = X_{N(1)}) \to 1$ as $n \to \infty$.*

Proof. Let $x_1, x_2, \ldots$ be the atoms of X, with masses $p_1, p_2, \ldots$. Fix a realization of $X_1, \ldots, X_n$. If $X_j \neq X_1$ for all $j \neq 1$, then $X_{n,1} = X_{N(1)}$. Suppose that $X_j = X_1$ for at least one $j \neq 1$. Let M be the number of such j. Then with probability $1/(M+1)$, $\pi(1)$ is the highest among all such $\pi(j)$. If this does not happen, then again $X_{n,1} = X_{N(1)}$. Therefore

$$\mathbb{P}(X_{n,1} \neq X_{N(1)}) \leq \mathbb{E}\biggl(\frac{1}{M+1} 1_{\{M \geq 1\}}\biggr).$$

Now let us condition on X_1. If $X_1 \notin \{x_1, x_2, \ldots\}$, then $M = 0$. If $X_1 = x_i$, then conditionally $M \sim \text{Binomial}(n-1, p_i)$. Therefore by Lemma A.7 and the above inequality, we get

$$\mathbb{P}(X_{n,1} \neq X_{N(1)}) \leq \sum_{i=1}^{\infty} \frac{1-(1-p_i)^n}{np_i} p_i.$$

Take any k. Then by the inequality $(1-x)^n \geq 1 - nx$ and the above inequality,

$$\mathbb{P}(X_{n,1} \neq X_{N(1)}) \leq \frac{k}{n} + \sum_{i=k+1}^{\infty} p_i.$$

Fixing k, and sending $n \to \infty$, we get

$$\limsup_{n\to\infty} \mathbb{P}(X_{n,1} \neq X_{N(1)}) \leq \sum_{i=k+1}^{\infty} p_i.$$

The proof is completed by sending $k \to \infty$. □

Corollary A.9. *For any measurable $f : \mathbb{R} \to \mathbb{R}$, $f(X_1) - f(X_{N(1)}) \to 0$ in probability as $n \to \infty$.*

Proof. By Lemma A.6, $f(X_1) - f(X_{n,1}) \to 0$ in probability. By Lemma A.8, $f(X_{n,1}) - f(X_{N(1)}) \to 0$ in probability. The claim is proved by adding the two. □

For each $t \in \mathbb{R}$, let

$$F_n(t) := \frac{1}{n} \sum_{i=1}^{n} 1_{\{Y_i \leq t\}}, \quad G_n(t) := \frac{1}{n} \sum_{i=1}^{n} 1_{\{Y_i \geq t\}}.$$

Define

$$Q_n := \frac{1}{n}\sum_{i=1}^n \min\{F_n(Y_i), F_n(Y_{N(i)})\} - \frac{1}{n}\sum_{i=1}^n G_n(Y_i)^2.$$

Lemma A.10. *Let Q_n be defined as above, and Q be the quantity defined in equation* (A.1)*. Then $\lim_{n\to\infty}\mathbb{E}(Q_n) = Q$.*

Proof. Let

$$Q_n' := \frac{1}{n}\sum_{i=1}^n \min\{F(Y_i), F(Y_{N(i)})\} - \frac{1}{n}\sum_{i=1}^n G(Y_i)^2.$$

and let

$$\Delta_n := \sup_{t\in\mathbb{R}} |F_n(t) - F(t)| + \sup_{t\in\mathbb{R}} |G_n(t) - G(t)|.$$

Then by the triangle inequality,

$$|Q_n' - Q_n| \le 3\Delta_n.$$

On the other hand, by the Glivenko–Cantelli theorem, $\Delta_n \to 0$ almost surely as $n\to\infty$. Since Δ_n is bounded by 2, this implies that

$$\lim_{n\to\infty} \mathbb{E}|Q_n' - Q_n| = 0.$$

Thus, it suffices to show that $\mathbb{E}(Q_n')$ converges to Q. First, notice that

$$\min\{F(Y_1), F(Y_{N(1)})\} = \int 1_{\{Y_1\ge t\}} 1_{\{Y_{N(1)}\ge t\}} d\mu(t).$$

Let $\mathcal{F}$ be the σ-algebra generated by the X_i's and the randomness used for breaking ties in the selection of π. Then for any t,

$$\mathbb{E}(1_{\{Y_1\ge t\}} 1_{\{Y_{N(1)}\ge t\}} | \mathcal{F}) = G_{X_1}(t) G_{X_{N(1)}}(t).$$

Now recall that by the properties of the regular conditional probability μ_x, the map $x \mapsto G_x(t)$ is measurable. Therefore by the above identity and Corollary A.9, and the boundedness of G_x, we have

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}(1_{\{Y_1\ge t\}} 1_{\{Y_{N(1)}\ge t\}}) &= \lim_{n\to\infty} \mathbb{E}(G_{X_1}(t) G_{X_{N(1)}}(t)) \\ &= \mathbb{E}(G_X(t)^2).\end{aligned}$$

Thus,

$$\lim_{n\to\infty} \mathbb{E}(Q_n') = \int_{\mathbb{R}} (\mathbb{E}(G_X(t)^2) - G(t)^2) d\mu(t).$$

Since $\mathbb{E}(G_X(t)) = G(t)$, this completes the proof of the lemma. □

Lemma A.11. *There is a positive universal constant C such that for any n and any $t \ge 0$,*

$$\mathbb{P}(|Q_n - \mathbb{E}(Q_n)| \ge t) \le 2e^{-Cnt^2}.$$

Proof. Throughout this proof, C will denote any universal constant. The value of C may change from line to line. First, we will prove the claim under the assumption that X has a continuous distribution, so that no randomization is involved in the definitions of π and the $N(i)$'s.

Assume continuity, and suppose that for some $i \le n$, (X_i, Y_i) is replaced by a different value (X_i', Y_i'). Then there are at most three indices j such that the value of $N(j)$ changes after the replacement, and exactly one index, $j = i$, where Y_j changes. Moreover, there can be at most one index j such that $N(j) = i$, both before and after the replacement. Lastly, for each t, $G_n(t)$ and $F_n(t)$ change by at most $1/n$. This shows that Q_n changes by at most C/n due to this replacement. The result now follows easily by the bounded difference concentration inequality [3].

Now consider the general case. Let $Z_1, \ldots, Z_n$ be i.i.d. Uniform$[0, 1]$ random variables. For each $\varepsilon > 0$, define

$$X_i^\varepsilon := X_i + \varepsilon Z_i.$$

Define Q_n^ε using $(X_1^\varepsilon, Y_1), \ldots, (X_n^\varepsilon, Y_n)$, by the same formula that was used for defining Q_n using $(X_1, Y_1), \ldots, (X_n, Y_n)$. Then by the first part we know that

$$\mathbb{P}(|Q_n^\varepsilon - \mathbb{E}(Q_n^\varepsilon)| \ge t) \le 2e^{-Cnt^2}, \tag{A.2}$$

where the important thing is that C has no dependence on ε. Now construct a random permutation π as follows. Given a realization of $X_1, \ldots, X_n$, let

$$\varepsilon^* := \frac{1}{2} \min\{|X_i - X_j| : 1 \le i, j \le n, X_i \ne X_j\}.$$

Having produced ε^* as above, define π to be the rank vector of $X_1^{\varepsilon^*}, \ldots, X_n^{\varepsilon^*}$. Notice that if $X_i < X_j$ for some i and j, then it is guaranteed that $X_i^{\varepsilon^*} < X_j^{\varepsilon^*}$. From this, it is not hard to see that π is a rank vector for $X_1, \ldots, X_n$ where ties are broken uniformly at random. On the other hand, the construction also guarantees that π is the rank vector $X_1^\varepsilon, \ldots, X_n^\varepsilon$ for all $\varepsilon \le \varepsilon^*$. Thus, if Q_n is defined using this π, then $Q_n^\varepsilon = Q_n$ for all $\varepsilon \le \varepsilon^*$. Consequently, $Q_n^\varepsilon \to Q_n$ almost surely as $\varepsilon \to 0$. Using the uniform boundedness of Q_n^ε, it is now easy to deduce the tail bound for Q_n from the inequality (A.2). □

Combining Lemmas A.10 and A.11, we get the following corollary.

Corollary A.12. *As $n \to \infty$, $Q_n \to Q$ almost surely.*

We are now ready to prove Theorem 1.1.

Proof of Theorem 1.1. Define

$$S_n := \frac{1}{n}\sum_{i=1}^n G_n(Y_i)(1 - G_n(Y_i)), \quad S_n' := \frac{1}{n}\sum_{i=1}^n G(Y_i)(1 - G(Y_i)), \tag{A.3}$$

and $\Delta_n := \sup_{t \in \mathbb{R}} |G_n(t) - G(t)|$. Then by the triangle inequality, $|S_n - S_n'| \le 2\Delta_n$, and by the Glivenko–Cantelli theorem, $\Delta_n \to 0$ almost surely. But by the strong law of large numbers, $S_n' \to \int G(t)(1-G(t))d\mu(t)$ almost surely as $n \to \infty$, and therefore the same holds for S_n. By Corollary A.2, this limit is nonzero. Therefore by this and Corollary A.12, we get that with probability one,

$$\lim_{n\to\infty} \frac{Q_n}{S_n} = \xi,$$

where $\xi = \xi(X,Y)$ is the quantity appearing the statement of Theorem 1.1. Now notice that if π is the permutation used for rearranging the data in the definition of ξ_n, then $nF_n(Y_i) = r_{\pi(i)}$ for all i, and $nF_n(Y_{N(i)}) = r_{\pi(i)+1}$ for $i \neq \pi^{-1}(n)$. If $i = \pi(n)$, then $nF_n(Y_i) = nF_n(Y_{N(i)}) = r_n$. Therefore

$$\frac{1}{n}\sum_{i=1}^n \min\{F_n(Y_i), F_n(Y_{N(i)})\} = \frac{1}{n^2}\sum_{i\neq \pi^{-1}(n)} \min\{r_{\pi(i)}, r_{\pi(i)+1}\} + \frac{r_n}{n^2}.$$

By the identity $\min\{a,b\} = \frac{1}{2}(a+b-|a-b|)$, this gives

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \min\{F_n(Y_i), F_n(Y_{N(i)})\}\\
&= \frac{1}{2n^2}\sum_{i\neq\pi^{-1}(n)} (r_{\pi(i)} + r_{\pi(i)+1} - |r_{\pi(i)} - r_{\pi(i)+1}|) + \frac{r_n}{n^2}\\
&= \frac{1}{n^2}\sum_{i=1}^n r_i - \frac{1}{2n^2}\sum_{i=1}^{n-1}|r_{i+1} - r_i| + \frac{r_n - r_1}{2n^2}.
\end{aligned}$$

On the other hand,

$$S_n = \frac{1}{n^3}\sum_{i=1}^n l_i(n-l_i), \quad \frac{1}{n}\sum_{i=1}^n G_n(Y_i)^2 = \frac{1}{n^3}\sum_{i=1}^n l_i^2,$$

and

$$\sum_{i=1}^n r_i = \sum_{i=1}^n\sum_{j=1}^n 1_{\{Y_{(j)}\le Y_{(i)}\}} = \sum_{j=1}^n\sum_{i=1}^n 1_{\{Y_{(j)}\le Y_{(i)}\}} = \sum_{j=1}^n l_j. \tag{A.4}$$

Combining the above observations, we get

$$\frac{Q_n}{S_n} = \xi_n + \frac{r_n - r_1}{2n^2 S_n}.$$

In particular,

$$\left|\frac{Q_n}{S_n} - \xi_n\right| \le \frac{1}{2nS_n}.$$

Since S_n converges to a nonzero limit, this proves that $\xi_n \to \xi$ almost surely. Since for each t,

$$G(t)(1-G(t)) = \mathrm{Var}(1_{\{Y\ge t\}}) \ge \mathrm{Var}(\mathbb{P}(Y\ge t|X)),$$

we conclude that $0 \le \xi \le 1$.

Lemma A.1 shows that $\xi = 0$ if and only if X and Y are independent. On the other hand, if Y is a function of X, say $Y = f(X)$ almost surely, then

$$\begin{aligned}\int \mathrm{Var}(\mathbb{P}(Y \ge t|X))d\mu(t) &= \int \mathrm{Var}(1_{\{f(X)\ge t\}})d\mu(t) \\ &= \int_{\mathbb{R}} \mathbb{P}(f(X) \ge t)(1 - \mathbb{P}(f(X) \ge t))d\mu(t) \\ &= \int G(t)(1 - G(t))d\mu(t),\end{aligned}$$

which shows that $\xi = 1$. Conversely, suppose that $\xi = 1$. Then by the law of total variance,

$$\begin{aligned}0 = 1 - \xi &= \int [\mathrm{Var}(1_{\{Y\ge t\}}) - \mathrm{Var}(\mathbb{P}(Y \ge t|X))]d\mu(t) \\ &= \int \mathbb{E}(\mathrm{Var}(1_{\{Y\ge t\}}|X))d\mu(t) \\ &= \int \mathbb{E}(G_X(t)(1 - G_X(t)))d\mu(t).\end{aligned}$$

This implies that $\mathbb{P}(E) = 1$, where E is the event

$$\int G_X(t)(1 - G_X(t))d\mu(t) = 0. \tag{A.5}$$

Let A be the support of μ. Define

$$a_x := \sup\{t : G_x(t) = 1\}, \quad b_x := \inf\{t : G_x(t) = 0\},$$

so that $a_x \le b_x$. By the measurability of $x \mapsto G_x(t)$ and the fact that $a_x \ge t$ if and only if $G_x(t) = 1$, it follows that $x \mapsto a_x$ is a measurable map. Similarly, $x \mapsto b_x$ is also measurable.

Now suppose that the event $\{a_X < b_X\} \cap E$ takes place. Since $G_X(t) \in (0, 1)$ for all $t \in (a_X, b_X)$, the condition (A.5) implies that $\mu((a_X, b_X)) = 0$. Since (a_X, b_X) is an open interval, this implies that $(a_X, b_X) \subseteq A^c$. On the other hand, under the given circumstance, we also have $\mathbb{P}(Y \in (a_X, b_X)|X) > 0$. Thus $\mathbb{P}(Y \in A^c|X) > 0$.

The above argument shows that if $\mathbb{P}(\{a_X < b_X\} \cap E) > 0$, then $\mathbb{P}(Y \in A^c) > 0$. But this is impossible, since A is the support of μ. Therefore $\mathbb{P}(\{a_X < b_X\} \cap E) = 0$. But $\mathbb{P}(E) = 1$. Therefore $\mathbb{P}(a_X = b_X) = 1$. Thus, $Y = a_X$ almost surely. This completes the proof of Theorem 1.1. □

B. Preparation for the proof of Theorem 2.2

In this section we prove some preparatory lemmas for the proof of Theorem 2.2. Let $R(i)$ be the number of j such that $Y_j \le Y_i$ and $L(i)$ be the number of j such that $Y_j \ge Y_i$. Let π be a rank vector for the X_i's, chosen uniformly at random from all available choices if there are ties. First, note that since X and Y are independent, π^{-1} is a uniform random permutation that is independent of $Y_1, \ldots, Y_n$. Let $\tau := \pi^{-1}$, and let

$$D_n := \sum_{i=1}^{n-1} a_i,$$

where

$$a_i := \min\{R(\tau(i)), R(\tau(i+1))\}.$$

Also, for convenience, let

$$b_{i,j} := \min\{R(i), R(j)\}.$$

In the following, $O(n^{-\alpha})$ will denote any quantity whose absolute value is bounded above by $Cn^{-\alpha}$ for some universal constant C. Let $\mathbb{E}'$, Var' and Cov' denote conditional expectation, conditional variance and conditional covariance given $Y_1, \ldots, Y_n$.

Lemma B.1.

$$\mathbb{E}'(D_n) = \frac{1}{n}\sum_{i=1}^{n} L(i)(L(i)-1).$$

Proof. Take any $1 \le i \le n-1$. Since $(\tau(i), \tau(i+1))$ is uniformly distributed over all pairs (j,k) where j and k are distinct, we have

$$\mathbb{E}'(a_i) = \frac{1}{n(n-1)} \sum_{1\le j\ne k\le n} b_{j,k} \tag{B.1}$$

Since $R(i) = \sum_{j=1}^n 1_{\{Y_j \le Y_i\}}$, this gives

$$\begin{aligned}
\mathbb{E}'(a_i) &= \frac{1}{n(n-1)} \sum_{1\le j\ne k\le n} \sum_{l=1}^{n} 1_{\{Y_l\le Y_j,\, Y_l\le Y_k\}} \\
&= \frac{1}{n(n-1)}\biggl(\sum_{1\le j,k\le n}\sum_{l=1}^{n} 1_{\{Y_l\le Y_j,\, Y_l\le Y_k\}} - \sum_{j=1}^{n}\sum_{l=1}^{n} 1_{\{Y_l\le Y_j\}}\biggr) \\
&= \frac{1}{n(n-1)}\biggl(\sum_{l=1}^{n}\sum_{1\le j,k\le n} 1_{\{Y_l\le Y_j,\, Y_l\le Y_k\}} - \sum_{l=1}^{n}\sum_{j=1}^{n} 1_{\{Y_l\le Y_j\}}\biggr) \\
&= \frac{1}{n(n-1)}\biggl(\sum_{l=1}^{n} L(l)^2 - \sum_{l=1}^{n} L(l)\biggr).
\end{aligned}$$

The proof is now completed by adding over i. □

Lemma B.2. $\mathrm{Var}'(D_n) = V_n + O(n^2)$, *where*

$$V_n := \frac{1}{n}\sum_{p,q=1}^{n} b_{p,q}^2 - \frac{2}{n^2}\sum_{p,q,r=1}^{n} b_{p,q}b_{p,r} + \frac{1}{n^3}\sum_{p,q,r,s=1}^{n} b_{p,q}b_{r,s}.$$

Proof. Take any $1 \le i < j \le n-1$. First, suppose that $i+1 < j$. Then $(\tau(i), \tau(i+1), \tau(j), \tau(j+1))$ is uniformly distributed over all quadruples of distinct (p,q,r,s). Thus,

$$\mathbb{E}'(a_i a_j) = \frac{1}{(n)_4} \sum_{p,q,r,s}{}' b_{p,q} b_{r,s},$$

where $(n)_4 := n(n-1)(n-2)(n-3)$, and $\sum'$ denotes sum over distinct p,q,r,s. Therefore by (B.1),

$$\begin{aligned}
\mathrm{Cov}'(a_i, a_j) &= \frac{1}{(n)_4} \sum_{p,q,r,s}{}' b_{p,q} b_{r,s} - \biggl(\frac{1}{(n)_2} \sum_{p,q}{}' b_{p,q}\biggr)^2 \\
&= \biggl(\frac{1}{(n)_4} - \frac{1}{(n)_2^2}\biggr) \sum_{p,q,r,s}{}' b_{p,q} b_{r,s} - \frac{1}{(n)_2^2}\biggl(\biggl(\sum_{p,q}{}' b_{p,q}\biggr)^2 - \sum_{p,q,r,s}{}' b_{p,q} b_{r,s}\biggr) \\
&= \frac{4n}{(n)_2 (n)_4} \sum_{p,q,r,s}{}' b_{p,q} b_{r,s} - \frac{4}{(n)_2^2} \sum_{p,q,r}{}' b_{p,q} b_{p,r} + O(1) \\
&= \frac{4}{n^5} \sum_{p,q,r,s} b_{p,q} b_{r,s} - \frac{4}{n^4} \sum_{p,q,r} b_{p,q} b_{p,r} + O(1).
\end{aligned}$$

Next, suppose that $i+1 = j$. Then

$$\begin{aligned}
\mathrm{Cov}'(a_i, a_j) &= \frac{1}{(n)_3} \sum_{p,q,r}{}' b_{p,q} b_{p,r} - \biggl(\frac{1}{(n)_2} \sum_{p,q}{}' b_{p,q}\biggr)^2 \\
&= \frac{1}{n^3} \sum_{p,q,r} b_{p,q} b_{p,r} - \frac{1}{n^4} \sum_{p,q,r,s} b_{p,q} b_{r,s} + O(n).
\end{aligned}$$

Similarly, if $i = j$, then

$$\begin{aligned}
\mathrm{Cov}'(a_i, a_j) &= \frac{1}{(n)_2} \sum_{p,q}{}' b_{p,q}^2 - \biggl(\frac{1}{(n)_2} \sum_{p,q}{}' b_{p,q}\biggr)^2 \\
&= \frac{1}{n^2} \sum_{p,q} b_{p,q}^2 - \frac{1}{n^4} \sum_{p,q,r,s} b_{p,q} b_{r,s} + O(n).
\end{aligned}$$

The proof is completed by adding up $\mathrm{Cov}'(a_i, a_j)$ over all $1 \le i, j \le n-1$. □

Lemma B.3. *As $n \to \infty$, $\mathrm{Var}'(D_n)/n^3$ converges almost surely to the deterministic limit*

$$\mathbb{E}(\phi(Y_1, Y_2)^2 - 2\phi(Y_1, Y_2)\phi(Y_1, Y_3) + \phi(Y_1, Y_2)\phi(Y_3, Y_4)),$$

where $\phi(y, y') := \min\{F(y), F(y')\}$ and Y_1, Y_2, Y_3, Y_4 are i.i.d. copies of Y.

Proof. Throughout this proof, C will be used to denote any universal constant. Let V_n be as in Lemma B.2. It is a function of the Y_i's only. Notice that if one Y_i is replaced by some other value Y_i', then each $R(j)$ changes by at most 1 for $j \neq i$, and $R(i)$ changes by at most n. Therefore $b_{p,q}$ changes by at most 1 if $p \neq i$ and $q \neq i$, and by at most n if one or both of the indices are equal to i. Moreover, the b_{pq}'s are all bounded by n. Thus, changing one Y_i to Y_i' changes V_n by at most Cn^2. Therefore by the bounded difference inequality,

$$\mathbb{P}(|V_n - \mathbb{E}(V_n)| \ge t) \le 2e^{-Ct^2/n^5}$$

for every t. Consequently, $(V_n - \mathbb{E}(V_n))/n^3 \to 0$ almost surely as $n \to \infty$.

On the other hand, note that $b_{p,q}/n = \min\{F_n(Y_p), F_n(Y_q)\}$, where F_n is the empirical distribution function of the Y_i's. By the Glivenko–Cantelli theorem, $F_n \to F$ uniformly with probability one, where F is the cumulative distribution function of Y. From this, it is easy to see that $\mathbb{E}(V_n)/n^3$ converges to the displayed limit. □

Lemma B.4. *If Y is not a constant, the limit in Lemma* B.3 *is strictly positive.*

Proof. Let us denote the limit by v. Let Y' be an independent copy of Y, and define

$$\psi(y) := \mathbb{E}(\phi(y, Y')) = \mathbb{E}(\phi(Y, Y')|Y = y).$$

Also, let $m := \mathbb{E}(\phi(Y, Y')) = \mathbb{E}(\psi(Y))$. Then v can be expressed as

$$v = \mathbb{E}(\phi(Y, Y')^2) - 2\mathbb{E}(\psi(Y)^2) + m^2. \tag{B.2}$$

Now,

$$\begin{aligned}
&\mathbb{E}(\phi(Y, Y') - \psi(Y) - \psi(Y') + m)^2 \\
&= \mathbb{E}(\phi(Y, Y')^2 + \psi(Y)^2 + \psi(Y')^2 + m^2 - 2\phi(Y, Y')\psi(Y) \\
&\quad - 2\phi(Y, Y')\psi(Y') + 2\phi(Y, Y')m + 2\psi(Y)\psi(Y') \\
&\quad - 2\psi(Y)m - 2\psi(Y')m).
\end{aligned}$$

Note that $\mathbb{E}(\phi(Y, Y')\psi(Y)) = \mathbb{E}(\psi(Y)^2)$, and recall that $\mathbb{E}(\phi(Y, Y')) = \mathbb{E}(\psi(Y)) = m$. The same identities hold if we exchange Y and Y'. Using these facts, it is now easy to verify that the above expression is actually equal to the right side of (B.2). Thus,

$$v = \mathbb{E}(\phi(Y, Y') - \psi(Y) - \psi(Y') + m)^2.$$

Hence $v \ge 0$, and $v = 0$ if and only if $\phi(Y, Y') = \psi(Y) + \psi(Y') - m$ almost surely. Suppose that this is true. Then almost surely for each $i \ge 2$,

$$\phi(Y_1, Y_i) = \psi(Y_1) + \psi(Y_i) - m, \tag{B.3}$$

where $Y_1, Y_2, \ldots$ are i.i.d. copies of Y. Taking the minimum over $2 \le i \le n$ on both sides, we get

$$\min\{F(Y_1), \ldots, F(Y_n)\} = \psi(Y_1) + \min\{\psi(Y_2), \ldots, \psi(Y_n)\} - m.$$

Now, the minimum of a sequence of i.i.d. bounded random variables converges almost surely to the infimum of the support. Also, F and ψ are bounded functions. Therefore taking $n \to \infty$ on both sides of the above, it follows that $\psi(Y_1)$ equals a constant almost surely. Therefore $\psi(Y_2)$ equals the same constant almost surely, and hence by (B.3), $\phi(Y_1, Y_2)$ is also equal to a constant almost surely. Now, if $L(t) := \mathbb{P}(F(Y) \ge t)$, then $\mathbb{P}(\phi(Y_1, Y_2) \ge t) = L(t)^2$. Since $\phi(Y_1, Y_2)$ is a constant, this shows that $L(t)^2$ is 0 or 1 for every t, and hence $L(t)$ is also 0 or 1 for every t. Consequently, $F(Y)$ is a constant almost surely.

We claim that 1 is in the support of $F(Y)$ and hence $F(Y) = 1$ almost surely. To see this, take any $\varepsilon \in (0, 1)$. We will show that $\mathbb{P}(F(Y) > 1 - \varepsilon) > 0$. Let $x := \inf\{y : F(y) \ge 1 - \varepsilon/2\}$. Then x is a finite real number since F tends to 1 at ∞ and to 0 at $-\infty$. By the right-continuity of F, $F(x) \ge 1 - \varepsilon/2$. If F is discontinuous at x, this immediately shows that $\mathbb{P}(F(Y) > 1 - \varepsilon) \ge \mathbb{P}(Y = x) > 0$. If F is continuous at x, there is some $y < x$ such that $F(y) > 1 - \varepsilon$. By the definition of x, $F(y) < F(x)$. Thus, $\mathbb{P}(F(Y) > 1 - \varepsilon) \ge \mathbb{P}(Y \in (y, x)) > 0$. This shows that 1 is in the support of $F(Y)$, and hence $F(Y) = 1$ almost surely.

Since Y is not a constant, there are at least two points in its support. Therefore there exist two disjoint nonempty open intervals I and J such that $\mathbb{P}(Y \in I)$ and $\mathbb{P}(Y \in J)$ are both positive. Suppose that I is to the left of J. Then for any $y \in I$, $F(y) \le 1 - \mathbb{P}(Y \in J) < 1$, and hence $\mathbb{P}(F(Y) < 1) \ge \mathbb{P}(Y \in I) > 0$, which contradicts the conclusion of the previous paragraph. This shows that $v > 0$. $\square$

C. Proof of Theorem 2.2

We will continue with the notations from Section B. Let σ^2 denote the limit of $\mathrm{Var}'(D_n)/n^3$, which by Lemmas B.3 and B.4, is a deterministic positive quantity (it was called v in the proof of Lemma B.4). Define

$$\widetilde{D}_n := \frac{D_n - \mathbb{E}'(D_n)}{n^{3/2}\sigma}.$$

Notice that $r_i = R(\tau(i))$. Therefore by Lemma B.1, the identity (A.4), and the identity $\min\{a, b\} = \frac{1}{2}(a + b - |a - b|)$, we get

$$\begin{aligned}
D_n - \mathbb{E}'(D_n) &= \sum_{i=1}^{n-1} \min\{r_i, r_{i+1}\} - \frac{1}{n}\sum_{i=1}^{n} L(i)(L(i) - 1) \\
&= \frac{1}{2}\sum_{i=1}^{n-1}(r_i + r_{i+1} - |r_{i+1} - r_i|) - \frac{1}{n}\sum_{i=1}^{n} l_i(l_i - 1) \\
&= \sum_{i=1}^{n} r_i - \frac{r_1 + r_n}{2} - \frac{1}{2}\sum_{i=1}^{n-1}|r_{i+1} - r_i| - \frac{1}{n}\sum_{i=1}^{n} l_i(l_i - 1) \\
&= \frac{1}{n}\sum_{i=1}^{n} l_i(n - l_i) - \frac{1}{2}\sum_{i=1}^{n-1}|r_{i+1} - r_i| + O(n).
\end{aligned}$$

This shows that

$$\xi_n = \frac{D_n - \mathbb{E}'(D_n)}{n^2 S_n} + O\biggl(\frac{1}{nS_n}\biggr) = \frac{\sigma}{\sqrt{n}S_n}\widetilde{D}_n + O\biggl(\frac{1}{nS_n}\biggr),$$

where S_n is the quantity defined in (A.3). In the proof of Theorem 1.1, we showed that $S_n \to \int G(t)(1-G(t)d\mu(t)$ almost surely, and the latter quantity is positive by Corollary A.2. Thus, to prove the central limit theorem for $\sqrt{n}\xi_n$, it suffices to prove the central limit theorem for $\widetilde{D}_n$. The formula for the limiting variance τ^2 can be read off from the limit of S_n and the formula for σ. The limiting variance is strictly positive by Lemma B.4. When Y is continuous, $F(Y) \sim$ Uniform$[0,1]$. Using this fact, an easy calculation shows that $\tau^2 = 2/5$.

The central limit theorem for $\widetilde{D}_n$ can be proved by mimicking the proof of the main theorem of the paper [1]. First, replace D_n by

$$D'_n := \sum_{i=1}^n \min\{R(\tau(i)), R(\tau(i+1))\},$$

where $\tau(n+1) := \tau(1)$. Since $|D'_n - D_n| \le n$, it suffices to prove that $\widetilde{D}'_n \to N(0,1)$ in distribution, where

$$\widetilde{D}'_n := \frac{D'_n - \mathbb{E}'(D'_n)}{n^{3/2}\sigma}.$$

Mimicking the main idea of [1], we define

$$f(\tau(i+1)) := \mathbb{E}'(\min\{R(\tau(i)), R(\tau(i+1))\}|\tau(i+1)),$$

and observe that

$$\mathbb{E}'(D'_n) = n\mathbb{E}[f(\tau(1))] = \sum_{i=1}^n f(i) = \sum_{i=1}^n f(\tau(i)).$$

Thus,

$$\widetilde{D}'_n = \frac{\sum_{i=1}^n \beta_i}{n^{3/2}\sigma},$$

where $\beta_i := \min\{R(\tau(i)), R(\tau(i+1))\} - f(\tau(i))$. Since $|D_n - D'_n| \le n$, $\mathrm{Var}'(D'_n)/n^3$ converges almost surely to σ^2. Using these observations, we can proceed exactly as in the proof of the main theorem of [1] to show that for every integer $k \ge 1$,

$$\mathbb{E}'[(\widetilde{D}'_n)^k] \to \mathbb{E}(Z^k) \text{ almost surely as } n \to \infty, \tag{C.1}$$

where $Z \sim N(0,1)$. On the other hand, a simple argument using the bounded difference inequality (viewing τ as the rank vector of i.i.d. random variables from any continuous distribution) shows that for any k,

$$\sup_{n\ge 1} \mathbb{E}|\widetilde{D}'_n|^k < \infty.$$

Therefore by (C.1) and uniform integrability, we conclude that for every integer $k \ge 1$,

$$\lim_{n\to\infty} \mathbb{E}[(\widetilde{D}'_n)^k] = \mathbb{E}(Z^k).$$

This completes the proof of Theorem 2.2.

D. Proof of Theorem 2.3

The quantity S_n define in (A.3) is the same as d_n, and in the proof of Theorem 1.1 we showed that S_n converges to the square-root of the denominator in the definition of τ^2. Recall the quantity V_n from Lemma B.2. By Lemma B.3, we know V_n/n^3 converges almost surely to the numerator in the definition of τ^2. We will now show that $a_n - 2b_n + c_n^2$ is the same as V_n/n^3.

From the definition of V_n, it is easy to see that the result will remain unchanged if we permute the $R(i)$'s and recompute V_n. So we can replace the $R(i)$'s by an increasing rearrangement $u_1, \ldots, u_n$. Redefine

$$b_{ij} := \min\{u_i, u_j\} = u_{\min\{i,j\}}.$$

Then it is clear that

$$\begin{aligned}\sum_{i,j} b_{ij} &= \sum_{i=1}^n u_i + 2 \sum_{1\le i<j\le n} b_{ij} \\ &= \sum_{i=1}^n u_i + 2\sum_{i=1}^n (n-i)u_i = \sum_{i=1}^n (2n-2i+1)u_i.\end{aligned}$$

Similarly,

$$\sum_{i,j} b_{ij}^2 = \sum_{i=1}^n (2n-2i+1)u_i^2.$$

Finally,

$$\begin{aligned}\sum_{i,j,k} b_{ij}b_{ik} &= \sum_{i=1}^n \biggl(\sum_{j=1}^n b_{ij}\biggr)^2 \\ &= \sum_{i=1}^n \biggl(\sum_{j=1}^i u_j + (n-i)u_i\biggr)^2 = \sum_{i=1}^n (v_i + (n-i)u_i)^2.\end{aligned}$$

These expressions make it clear that $a_n - 2b_n + c_n^2 = V_n/n^3$. This completes the proof of convergence. Finally, to see that $\widehat{\tau}_n^2$ can be computed in time $O(n \log n)$, simply observe that the computation involves only sorting and calculating cumulative sums, both of which can be done in time $O(n \log n)$.

References

[1] Chao, C.-C., Bai, Z. and Liang, W.-Q. (1993). Asymptotic normality for oscillation of permutation. *Probab. Eng. Inf. Sci.,* **7**, 227–235.

[2] Durrett, R. (2010). *Probability: theory and examples.* Fourth edition. Cambridge University Press, Cambridge.

[3] McDiarmid, C. (1989). On the method of bounded differences. In *Surveys in combinatorics,* 148–188, London Math. Soc. Lecture Note Ser., 141, Cambridge Univ. Press, Cambridge.

[4] Rudin, W. (1987). *Real and complex analysis.* Third edition. McGraw-Hill Book Co., New York.

Department of Statistics, Stanford University, Sequoia Hall, 390 Jane Stanford Way, Stanford, CA 94305
Email address: souravc@stanford.edu