The Support of the Distinct-Tuple Distribution

2026-05-01

1 At a glance

This note concerns the distinct-tuple distribution used in the pasting section of [JNV⁺20].[1]

- **Difficulty.** Low. The issue is a boundary convention for an empty finite support, not a change to the proof of the pasting theorem.
- **Estimated weight.** Small. The formal repair is already in place; this note records how the all-k Lean object agrees with the paper's probability-language statement when $k \le q$.
- **Mathlib/project split.** Mostly project-local. The argument uses finite supports and total variation estimates from the project distribution API; Mathlib supplies the finite-type background.
- **Key mathematical inputs.** The set $\mathsf{Distinct}_k \subseteq \mathbb{F}_q^k$ is nonempty exactly when $k \le q$; when $k > q$, the displayed k^2/q error term is already large enough to make the comparison trivial after the formal object is interpreted as the zero finite weight.

2 Key theorem forms

For $k \ge 1$, the paper defines

$$\mathsf{Distinct}_k = \{(x_1, \ldots, x_k) \in \mathbb{F}_q^k : x_i \neq x_j \text{ for } i \neq j\}$$

[1] Tracked as part of the discrepancy audit in GitHub issue #930. In the TeX source consulted here, the definition and comparison proposition are in `references/ldt-paper/ld-pasting.tex`, lines 167–213.

and writes $(\mathbf{y}_1, \ldots, \mathbf{y}_k) \sim \mathsf{Distinct}_k$ for a uniformly random element of this set. The following comparison proposition then bounds the total variation distance between a uniformly random tuple in $\mathbb{F}_q^k$ and a uniformly random tuple in $\mathsf{Distinct}_k$ by k^2/q.

3 The point at issue

The distribution on $\mathsf{Distinct}_k$ is a genuine probability distribution only when $\mathsf{Distinct}_k$ is nonempty. This requires $k \le q$. The statement of the comparison proposition and the surrounding pasting theorem do not state this condition. Later in the section the paper assumes $k \ge 400md$, but it does not separately assume $k \le q$.

Thus, read literally, the phrase "uniformly random element of $\mathsf{Distinct}_k$" is undefined in the case $k > q$. In that case no injective map from $\{1, \ldots, k\}$ to $\mathbb{F}_q$ exists. The displayed bound k^2/q is then already at least 1, so the intended comparison is only used in a regime where the estimate is trivial, but the probability distribution itself still needs a formal interpretation.

4 The blueprint statement

The blueprint repeats the paper's convention rather than making the boundary case explicit. The node `def:distinct-tuples` in `blueprint/src/chapter/ch09_pasting.tex`, lines 198–202, carries `\lean` annotations for MIPStarRE.LDT.Pasting.distinctTuples and MIPStarRE.LDT.Pasting.distinctTupleDistribution, together with `\leanok`, and states: "For $k \ge 1$, let $\mathsf{Distinct}_k \subseteq \mathbb{F}_q^k$ be the set of tuples $(x_1, \ldots, x_k)$ with pairwise distinct coordinates." This is the same all-$k \ge 1$ convention as the paper; the blueprint definition does not add the missing nonemptiness hypothesis $k \le q$.

The companion node `prop:ld-dnoteq` in the same file, lines 204–226, has a `\lean` annotation for MIPStarRE.LDT.Pasting.ldDnoteq and is also tagged `\leanok`. It states that if $x = (x_1, \ldots, x_k)$ is uniformly random in $\mathbb{F}_q^k$ and $y = (y_1, \ldots, y_k)$ is uniformly random in $\mathsf{Distinct}_k$, then $d_{\mathrm{TV}}(x, y) \le k^2/q$. Again, the blueprint records no $k \le q$ side condition for the phrase "uniformly random in $\mathsf{Distinct}_k$".

Thus the Lean repair is invisible from the blueprint side: a reader following the formalized blueprint nodes sees the same probability-language statement as in the paper, while the checked Lean declarations below implement the total object as a finite nonnegative weight for all k and recover a probability interpretation only when the support is nonempty, equivalently when $k \le q$.

5 The formal repair

The formal development uses a finite-support weighted distribution rather than forcing every distribution object to be normalized. The declaration MIPStarRE.LDT.Pasting.distinctTupleDistribution assigns weight $1/|\mathsf{Distinct}_k|$ to each distinct tuple when the support is nonempty, and weight zero outside the support. If $k > q$, the support is empty, so this is the zero finite measure rather than a probability distribution.[2]

The formalization records both facts. The general theorem MIPStarRE.LDT.Pasting.distinctTupleDistribution_weight_sum_le_one proves that the total mass is at most 1 for every k. The normalized form MIPStarRE.LDT.Pasting.distinctTupleDistribution_weight_sum_eq_one_of_le adds the missing hypothesis $k \le q$ and proves that the total mass is exactly 1.[3]

The total-variation comparison MIPStarRE.LDT.Pasting.ldDnoteq is correspondingly case-split. When $k \le q$, it proves the usual collision-probability estimate. When $k > q$, the support is empty; the formal total variation between the uniform tuple distribution and the zero distinct-tuple weight is $1/2$, and this is bounded by k^2/q because $k > q$.[4] Downstream averaging lemmas use the same convention: they compare expectations of nonnegative functions bounded by 1, and the $k > q$ branch is discharged by the now-trivial k^2/q error term.

6 Conclusion

The paper implicitly treats the distinct-tuple distribution as if it were available for all $k \ge 1$. A literal probability distribution exists only for $k \le q$. The formal development makes the boundary case total by interpreting the distinct-tuple object as a finite nonnegative weight for all k, and as a probability distribution precisely under the additional hypothesis $k \le q$. This is a harmless formal repair: it preserves the paper argument in the nonempty-support case and makes the empty-support case a trivial bound rather than an undefined sampling operation.

[2] See `MIPStarRE/LDT/Pasting/Defs/Tuples.lean`, lines 23–36.

[3] The normalized statement is in `MIPStarRE/LDT/Pasting/ComparisonLemmas/Common.lean`, lines 145–178 of the source file.

[4] See `MIPStarRE/LDT/Pasting/Core.lean`, lines 561–805.

References

[JNV+20] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. Quantum soundness of the classical low individual degree test, 2020. arXiv:2009.12982.