We thank the reviewers for their professional work, constructive and thorough criticism, and sensible directions for improvement. Altogether, their judgement suggests an acceptance. Below we elaborate a response on the main criticisms that we will incorporate on strengthened final version.

Typos and lack of consistency/clarity:

- **Reviewer 2** pointed to typos in lines 136, 140, 166, lack of clarity in the definition of Z_g and lack of clarity in the definition of the entropic cost in terms of a generic cost function $c(x, y)$. We fully agree with the criticism and will improve the manuscript based on it.
- **Reviewer 2** points to an inconsistency in Figure 1. There, both ϵ and σ_g^2 can be used in the legend but we will stick to one to keep notation simple. She also suggests introducing $\mathcal{F}_\sigma$ earlier to improve readability, and we will do so.
- **Reviewer 2 Reviewer 4** points there is a reference to a bound for EL^2 that is not stated elsewhere. There, we meant EL and the bound is the one of Proposition 2, as **Reviewer 4** guessed.
- **Reviewer 4** points to a missing (b) in page 6. Indeed, (c) should have been (b). We will correct this.
- **Reviewer 4** points it is not clear that (f, g) in Theorem 3 are the optimal potentials. We will make clear whenever we write (f, g) we are referring to them.

Hypothesis: Reviewer 2 suggest being finer in stating which hypothesis are required for each of our improvements. Indeed, subgaussianity is needed for the extension to unbounded domains, but the exponential factor improvement is a separate contribution that would hold under the same hyphotesis of Genevay et al. (2019). We will make this clear.

Error in proof of theorem 3: We thank **Reviewer 4** for pointing out. We do agree there is an error in the proof, but it can be easily fixed. It can be shown that as long as β is σ-finite it is still possible to make sense of the relative entropy as if β was a probability measure. This problem has been addressed by e.g. Léonard (2014): essentially, we could simply define the relative entropy $H(\alpha|\beta)$ for unbounded β by the usual formula (line 55), and then what remains to verify is that we can still go from the definition in equation (14) to equations (15) and (16). In Léonard (2014) it is argued all measure theoretical subtleties can be sorted out. We will fix the prove by comment on this.

Lack of distributional limits: **Reviewer 4** pointed out a weakness of our CLT is that it expresses in terms of the centering constants $E(S(P_n, Q_n))$ instead of $S(P, Q)$. We agree it is a weakness, and we don't know if this can be improved, which boils down to answering whether the sample complexity $O(1/\sqrt{n})$ could be indeed $o(1/\sqrt{n})$. We will comment on this issue. Nonetheless, in practice it is observed that regardless of the rate in n, dependence on dimension can be bad (indeed, exponential, as shown in Goldfeld et al. (2019) for entropy estimation), making the construction of confidence intervals hard, even if the CLT holds with $S(P, Q)$ instead of $E(S(P_n, Q_n))$.

Other entropy estimators: **Reviewer 4** points to the recent work of Berrett et al. (2019) showing $\sqrt{n}$ consistent estimators for a certain class of smooth densities, with available distributional limits. We will cite this relevant work, but it isn't clear Berrett et al. (2019) is superseding our method, as for the CLT of Berrett et al. (2019) to kick in, it is required that a certain parameter that the number of neighbors k is greater than a constant that may depends wildly on dimension. It is shown (empirically) in Goldfeld et al. (2019) that their estimator (and hence, ours) outperforms the one of Berrett et al. (2019) in practice. Finally, it is unclear whether the method of Berrett et al. (2019) can be applied to all distributions in our setup; only for compactly supported P the inclusion is straightforward.

Comments on lower bounds, optimal constant as a function of subgaussian constant, heavier tailed distributions: **Reviewers 1 and 4** mentioned several improvement directions related to sharper statements of bounds, and extending them to heavier tailed distributions. We recognize them as valuable directions, and at least we will comment on these issues. However, we don't compromise for a definite answer of this issues but rather defer them as future work. In particular, we will try to provide lower bounds, e.g. arguing as in Goldfeld et al. (2019) and explore other cases beyond subgaussianity. Empirical results suggest they are valid in heavier tailed cases as well, but the extension is nontrivial because our machinery relies heavily in subgaussianity.

References

Berrett, T. B., Samworth, R. J., Yuan, M., et al. (2019). Efficient multivariate entropy estimation via k-nearest neighbour distances. *The Annals of Statistics*, 47(1):288–318.

Genevay, A., Chizat, L., Bach, F., , Cuturi, M., and Peyré, G. (2019). Sample complexity of sinkhorn divergences. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS).*

Goldfeld, Z., Greenewald, K., Polyanskiy, Y., and Weed, J. (2019). Convergence of smoothed empirical measures with applications to entropy estimation. *arXiv preprint arXiv:1905.13576.*

Léonard, C. (2014). Some properties of path measures. In *Séminaire de Probabilités XLVI*, pages 207–230. Springer.