

impl TopKMeasure for MaxDivergence

Tudor Cebere

Michael Shoemate

August 20, 2025

This proof resides in “**contrib**” because it has not completed the vetting process.

This document proves soundness of `permute_and_flip` [2] in `mod.rs` at commit `e62b0aa2` (outdated¹). `permute_and_flip` noisily selects the index of the greatest score from a vector of input scores.

Permute and flip is equivalent to report noisy max with exponential noise [1]. Report noisy max exponential is implemented via permute and flip because of its discrete nature. Implementation-wise, we will follow permute-and-flip, yet prove the correctness of the algorithm via this equivalence.

1 Hoare Triple

Precondition

Compiler-verified

- Method `noisy_top_k` *Types consistent with pseudocode.*
- Method `privacy_map` *Types consistent with pseudocode.*

Caller-verified

- Method `noisy_top_k`
 - `x` elements are non-null.
 - `scale` is finite and non-negative.
- Method `privacy_map`
 - `d_in` is non-null and positive.
 - `scale` is non-null and positive.

Pseudocode

```
1 class MaxDivergence(TopKMeasure):
2     @staticmethod
3     def noisy_top_k(x: list[TIA], scale: f64, k: usize, negate: bool) -> list[usize]:
4         return exponential_top_k(x, scale, k, negate)
5
6     @staticmethod
7     def privacy_map(d_in: f64, scale: f64) -> f64:
8         return d_in.inf_div(scale)
```

¹See new changes with `git diff e62b0aa2..865aa008 rust/src/measurements/noisy_top_k/mod.rs`

Postcondition

Theorem 1.1. The implementation is consistent with all associated items in the **TopKMeasure** trait.

1. Method **noisy_top_k**:

- Returns the index of the top element z_i , where each $z_i \sim \text{DISTRIBUTION}(\text{shift} = y_i, \text{scale} = \text{scale})$, and each $y_i = -x_i$ if **negate**, else $y_i = x_i$, k times with removal.
- Errors are data-independent, except for exhaustion of entropy.

2. Method **privacy_map**: For any x, x' where $d_{\text{in}} \geq d_{\text{Range}}(x, x')$, return $d_{\text{out}} \geq D_{\text{self}}(f(x), f(x'))$, where $f(x) = \text{noisy_top_k}(x = x, k = 1, \text{scale} = \text{scale})$.

Definition 1.2. A random variable follows the Exponential distribution if it has density

$$f(x) = \frac{1}{\beta} e^{-z} \quad (1)$$

where $z = \frac{x-\mu}{\beta}$, μ is the shift (location) parameter and β is the scale parameter.

*Proof of postcondition: **noisy_top_k**.* The preconditions of **exponential_noisy_max** are met, therefore by the postcondition of **exponential_top_k**, the postcondition of **noisy_top_k** is satisfied. \square

Before proving the privacy guarantees, we state a few required definitions and lemmas:

Definition 1.3. Report noisy max with exponential noise computes the index of the maximum element from a set of candidates $u \in \mathbf{d}_{\text{in}}$, adds isotropic exponential noise $Z_i \sim \text{Exp}(1/\lambda)$ to each element in the candidate set u and returns the maximum index as follows:

$$\text{RNM-Exp}(s) = \text{argmax}_i(s_i + Z_i), Z_i \sim \text{Exp}(\lambda) \quad (2)$$

Lemma 1.4. The permute-and-flip mechanism is equivalent to the report-noisy-max with exponential noise mechanism.

See [1] for proof of Lemma 1.4.

Lemma 1.5. Let $X_1, X_2 \sim \text{Exp}(\lambda)$, $\Delta \geq 0$, then

$$\Pr[X_1 - X_2 \geq \Delta] = e^{-\Delta\lambda} \Pr[X_1 - X_2 \geq 0] \quad (3)$$

Proof of Lemma 1.5.

$$\Pr[X_1 - X_2 \geq \Delta] \quad (4)$$

$$= 1 - \Pr[X_1 \leq \Delta + X_2] \quad (5)$$

$$= 1 - \int_0^\infty \Pr[X_1 \leq \Delta + X_2 | X_2 = x] \Pr[X_2 = x] dx \quad \text{by Law of Total Probability} \quad (6)$$

$$= 1 - \int_0^\infty \Pr[X_1 \leq \Delta + x] \Pr[X_2 = x] dx \quad \text{by the fact that } \Delta > 0 \quad (7)$$

$$= 1 - \int_0^\infty \lambda(1 - e^{-(x+\Delta)\lambda}) e^{-x\lambda} dx \quad (8)$$

$$= 1 - \lambda \int_0^\infty e^{-x\lambda} dx + \lambda e^{-\Delta\lambda} \int_0^\infty e^{-2x\lambda} dx \quad (9)$$

$$= 1 - 1 + e^{-\Delta\lambda}/2 \quad \Pr[X_1 - X_2 \leq 0] = \Pr[X_1 - X_2 \geq 0] = 1/2 \quad (10)$$

$$= e^{-\Delta\lambda} \Pr[X_1 - X_2 \geq 0] \quad (11)$$

\square

Lemma 1.6. Let $u, v \in \text{input_domain}$ be two vectors of scores. Assume u, v in input_domain are $\mathbf{d_in}$ -close under **LInfDistance** and $\text{privacy_map}(\mathbf{d_in}) \leq \mathbf{d_out}$. Let $Z^* = \min_{Z_i} \{u_i + Z_i \geq u_j + Z_j\}, \forall i \neq j$. Then

$$\ln \left(\frac{\Pr[\text{function}(u) = i]}{\Pr[\text{function}(v) = i]} \right) = \ln \left(\frac{\Pr[Z_i \geq Z^*]}{\Pr[Z_i \geq Z^* + \mathbf{d_in}]} \right) \quad (12)$$

Proof.

$$\ln \left(\frac{\Pr[\text{function}(u) = i]}{\Pr[\text{function}(v) = i]} \right) \quad (13)$$

$$= \ln \left(\frac{\Pr[\text{RNM-Exp}(u) = i]}{\Pr[\text{RNM-Exp}(v) = i]} \right) \quad \text{by Lemma 1.4} \quad (14)$$

$$= \ln \left(\frac{\Pr[\arg\max_k (u_k + Z_k) = i]}{\Pr[\arg\max_k (v_k + Z_k) = i]} \right) \quad \text{by Definition 1.3} \quad (15)$$

Observe that for a fixed i , report noisy max outputs i if:

$$u_i + Z^* \geq u_j + Z_j, \forall i \neq j \quad \Longleftrightarrow \quad (16)$$

$$u_i + (v_i - v_i) + Z^* \geq u_j + (v_j - v_j) + Z_j \quad \Longleftrightarrow \quad (17)$$

$$v_i + (u_i - v_i) + Z^* \geq v_j + (u_j - v_j) + Z_j \quad \Longleftrightarrow \quad (18)$$

$$v_i + ((u_i - v_i) - (u_j - v_j) + Z^*) \geq v_j + Z_j \quad \Longleftrightarrow \quad (19)$$

$$v_i + (\Delta + Z^*) \geq v_j + Z_j \quad (20)$$

In other words, if $Z_i \geq (\Delta + Z^*)$, then $\text{function}(u) = \text{function}(v) = i$. This yields us:

$$\ln \left(\frac{\Pr[\arg\max_k (u_k + Z_k) = i]}{\Pr[\arg\max_k (v_k + Z_k) = i]} \right) = \ln \left(\frac{\Pr[Z_i \geq Z^*]}{\Pr[Z_i \geq \Delta + Z^*]} \right) \quad (21)$$

□

*Proof of postcondition: **privacy_map**.*

$$\max_{u \sim v} D_\infty(M(u) | M(v)) \quad (22)$$

$$= \max_{u \sim v} \max_i \ln \left(\frac{\Pr[\text{function}(u) = i]}{\Pr[\text{function}(v) = i]} \right) \quad (23)$$

$$= \max_{u \sim v} \max_i \ln \left(\frac{\Pr[\text{RNM-Exp}(u) = i]}{\Pr[\text{RNM-Exp}(v) = i]} \right) \quad \text{by Lemma 1.4} \quad (24)$$

$$= \max_{u \sim v} \max_i \ln \left(\frac{\Pr[\arg\max_k (u_k + Z_k) = i]}{\Pr[\arg\max_k (v_k + Z_k) = i]} \right) \quad \text{by Definition 1.3} \quad (25)$$

$$= \max_{u \sim v} \max_i \ln \left(\frac{\Pr[Z_i \geq Z^*]}{\Pr[Z_i \geq Z^* + \mathbf{d_in}]} \right) \quad \text{by Lemma 1.6} \quad (26)$$

$$\leq \frac{\mathbf{d_in}}{\text{scale}} \quad \text{by Lemma 1.5} \quad (27)$$

□

References

- [1] Zeyu Ding, Daniel Kifer, Thomas Steinke, Yuxin Wang, Yingtai Xiao, Danfeng Zhang, et al. The permute-and-flip mechanism is identical to report-noisy-max with exponential noise. *arXiv preprint arXiv:2105.07260*, 2021.

- [2] Ryan McKenna and Daniel Sheldon. Permute-and-flip: A new mechanism for differentially private selection, 2020.