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This document proves soundness of `permute_and_flip` [2] in `mod.rs` at commit e62b0aa2 (outdated[1]). `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

```python
class MaxDivergence(TopKMeasure):
    @staticmethod
    def noisy_top_k(x: list[TIA], scale: f64, k: usize, negate: bool) -> list[usize]:
        return exponential_top_k(x, scale, k, negate)

    @staticmethod
    def privacy_map(d_in: f64, scale: f64) -> f64:
        return d_in.inf_div(scale)
```

[1] See new changes with `git diff e62b0aa2..34f38c0d 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} = $ `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

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

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. □

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 $u \in$ `d_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:

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

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

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

Proof of Lemma 1.5.

$$
\begin{aligned}
&\Pr[X_1 - X_2 \geq \Delta] && (4)\\
&= 1 - \Pr[X_1 \leq \Delta + X_2] && (5)\\
&= 1 - \int_0^\infty \Pr[X_1 \leq \Delta + X_2 | X_2 = x] \Pr[X_2 = x] dx \qquad \text{by Law of Total Probability} && (6)\\
&= 1 - \int_0^\infty \Pr[X_1 \leq \Delta + x] \Pr[X_2 = x] dx \qquad \text{by the fact that } \Delta > 0 && (7)\\
&= 1 - \int_0^\infty \lambda(1 - e^{-(x+\Delta)\lambda}) e^{-x\lambda} dx && (8)\\
&= 1 - \lambda \int_0^\infty e^{-x\lambda} dx + \lambda e^{-\Delta\lambda} \int_0^\infty e^{-2x\lambda} dx && (9)\\
&= 1 - 1 + e^{-\Delta\lambda}/2 \qquad \Pr[X_1 - X_2 \leq 0] = \Pr[X_1 - X_2 \geq 0] = 1/2 && (10)\\
&= e^{-\Delta\lambda} \Pr[X_1 - X_2 \geq 0] && (11)
\end{aligned}
$$

□

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

$$\ln\left(\frac{\Pr[\texttt{function}(u) = i]}{\Pr[\texttt{function}(v) = i]}\right) = \ln\left(\frac{\Pr[Z_i \geq Z^*]}{\Pr[Z_i \geq Z^* + \texttt{d\_in}]}\right) \tag{12}$$

Proof.

$$\begin{aligned}
&\ln\left(\frac{\Pr[\texttt{function}(u) = i]}{\Pr[\texttt{function}(v) = i]}\right) && && (13)\\
&= \ln\left(\frac{\Pr[\texttt{RNM-Exp(u)} = i]}{\Pr[\texttt{RNM-Exp}(v) = i]}\right) && \text{by Lemma 1.4} && (14)\\
&= \ln\left(\frac{\Pr[\text{argmax}_k(u_k + Z_k) = i]}{\Pr[\text{argmax}_k(v_k + Z_k) = i]}\right) && \text{by Definition 1.3} && (15)
\end{aligned}$$

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

$$\begin{aligned}
&u_i + Z^* \geq u_j + Z_j, \forall i \neq j && \iff && (16)\\
&u_i + (v_i - v_i) + Z^* \geq u_j + (v_j - v_j) + Z_j && \iff && (17)\\
&v_i + (u_i - v_i) + Z^* \geq v_j + (u_j - v_j) + Z_j && \iff && (18)\\
&v_i + ((u_i - v_i) - (u_j - v_j) + Z^*) \geq v_j + Z_j && \iff && (19)\\
&v_i + (\Delta + Z^*) \geq v_j + Z_j && && (20)
\end{aligned}$$

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

$$\ln\left(\frac{\Pr[\text{argmax}_k(u_k + Z_k) = i]}{\Pr[\text{argmax}_k(v_k + Z_k) = i]}\right) = \ln\left(\frac{\Pr[Z_i \geq Z^*]}{\Pr[Z_i \geq \Delta + Z^*]}\right) \tag{21}$$

□

Proof of postcondition: `privacy_map`.

$$\begin{aligned}
&\max_{u \sim v} D_\infty(M(u)|M(v)) && && (22)\\
&= \max_{u \sim v} max_i \ln\left(\frac{\Pr[function(u) = i]}{\Pr[function(v) = i]}\right) && && (23)\\
&= \max_{u \sim v} max_i \ln\left(\frac{\Pr[\texttt{RNM-Exp(u)} = i]}{\Pr[\texttt{RNM-Exp}(v) = i]}\right) && \text{by Lemma 1.4} && (24)\\
&= \max_{u \sim v} max_i \ln\left(\frac{\Pr[\text{argmax}_k(u_k + Z_k) = i]}{\Pr[\text{argmax}_k(v_k + Z_k) = i]}\right) && \text{by Definition 1.3} && (25)\\
&= \max_{u \sim v} max_i \ln\left(\frac{\Pr[Z_i \geq Z^*]}{\Pr[Z_i \geq Z^* + \texttt{d_in}]}\right) && \text{by Lemma 1.6} && (26)\\
&\leq \frac{\texttt{d_in}}{\texttt{scale}} && \text{by Lemma 1.5} && (27)
\end{aligned}$$

□

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.