fn discrete_gaussian_scale_to_accuracy

Michael Shoemate

April 15, 2024

This document contains materials associated with `discrete_gaussian_scale_to_accuracy`.

Definition 0.1. Let z be the true value of the statistic and X be the random variable the noisy release is drawn from. Define $Y = |X - z|$, the distribution of DP errors. Then for any statistical significance level `alpha`, denoted $\alpha \in [0, 1]$, and `accuracy`, denoted $a \geq 0$,

$$\alpha = P[Y \geq a] \tag{1}$$

Theorem 0.2. For any `scale` ≥ 0 denoted s, when $X \sim \mathcal{N}_{\mathbb{Z}}(z, s)$,

$$a = argmin_i \left[(1 - \alpha) \cdot \sum_{y \in \mathbb{Z}} e^{-(y/s)^2/2} \leq \sum_{x=0}^{i-1} (1 + 1[x \neq 0]) e^{-(x/s)^2/2} \right] \tag{2}$$

That is, the accuracy is the smallest i such that the inequality holds.

Proof. Consider that the distribution of $(X - z) \sim \mathcal{N}_{\mathbb{Z}}(0, s)$. Then the PMF of Y is:

$$\forall y \geq 0 \qquad g(y) = \frac{(1 + 1[y \neq 0]) e^{-(y/s)^2/2}}{\sum_{y \in \mathbb{Z}} e^{-(y/s)^2/2}} \tag{3}$$

The purpose of the indicator function is to avoid double-counting zero.
Now derive an expression for α:

$$\begin{aligned}
\alpha &= P[Y \geq a] \\
&= 1 - P[Y < a] \\
&= 1 - \sum_{y=0}^{a-1} g(y) && \text{where } g(y) \text{ is the distribution of Y} \\
&= 1 - \sum_{y=0}^{a-1} \frac{(1 + 1[y \neq 0]) e^{-(y/s)^2/2}}{\sum_{z \in \mathbb{Z}} e^{-(z/s)^2/2}}
\end{aligned}$$

Reorder terms:

$$(1 - \alpha) \sum_{z \in \mathbb{Z}} e^{-(z/s)^2/2} = \sum_{y=0}^{a-1} (1 + 1[y \neq 0]) e^{-(y/s)^2/2}$$

The accuracy is the smallest a for which the right term is greater than or equal to the left term. □

1 Implementation

The discrete bound only differs significantly from the continuous bound when the scale is small. When the scale is small, the terms approach zero relatively quickly, due to the exponential term. Since the probability mass away from the origin is monotonically decreasing, the left-hand side of the equation can be approximated (down to float error) by summing the masses until underflow.

We then simply run a linear search for the smallest a such that the inequality holds.