fn sample_discrete_gaussian

Michael Shoemate

February 20, 2025

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

Proves soundness of `fn sample_discrete_gaussian` in `mod.rs` at commit 0be3ab3e6 (outdated[1]). This proof is an adaptation of subsection 5.3 of [CKS20].

Vetting history

- Pull Request #519

1 Hoare Triple

Precondition

`scale` $\in \mathbb{Q} \land$ `scale` ≥ 0

Implementation

```python
def sample_discrete_gaussian(scale) -> int:
    if scale == 0:
        return 0

    t = floor(scale) + 1 #
    sigma2 = scale**2

    while True:
        candidate = sample_discrete_laplace(t) #
        x = abs(candidate) - sigma2 / t
        bias = x**2 / (2 * sigma2)  #
        if sample_bernoulli_exp(bias): #
            return candidate
```

Postcondition

For any setting of the input parameter `scale` such that the given preconditions hold, `sample_discrete_gaussian` either returns `Err(e)` due to a lack of system entropy, or `Ok(out)`, where `out` is distributed as $\mathcal{N}_{\mathbb{Z}}(0, scale^2)$.

[1] See new changes with `git diff 0be3ab3e6..71801603 rust/src/traits/samplers/cks20/mod.rs`

2 Proof

Definition 2.1. (Discrete Gaussian). [CKS20] Let $\mu, \sigma \in \mathbb{R}$ with $\sigma > 0$. The discrete gaussian distribution with location μ and scale σ is denoted $\mathcal{N}_{\mathbb{Z}}(\mu, \sigma^2)$. It is a probability distribution supported on the integers and defined by

$$\forall x \in \mathbb{Z} \quad P[X = x] = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sum_{y \in \mathbb{Z}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}} \quad \text{where } X \sim \mathcal{N}_{\mathbb{Z}}(\mu, \sigma^2)$$

Lemma 2.2. `sample_discrete_gaussian` only returns `Err(e)` when there is a lack of system entropy.

Proof. By the non-negativity precondition on `scale`, `t` on line 5 is non-negative, so the precondition on `sample_discrete_laplace` is met. Similarly, since `bias` on line 11 is non-negative, the preconditions on `sample_bernoulli_exp` are met. By the definitions of `sample_discrete_laplace` and `sample_bernoulli_exp`, an error is only returned when there is a lack of system entropy. The only source of errors in `sample_discrete_gaussian` is from the invocation of these functions, therefore `sample_discrete_gaussian` only returns `Err(e)` when there is a lack of system entropy. □

We now condition on not returning an error. Let $t = \lfloor \sigma \rfloor + 1$, and fix any iteration of the loop.

Lemma 2.3. [CKS20] If y is a realization of $Y \sim \mathcal{L}_{\mathbb{Z}}(0, t)$, and c is a realization of $C \sim Bernoulli(exp(-(|y| - \sigma^2/t)^2/(2\sigma^2)))$, then $E[C] = \frac{1-e^{-1/\sigma}}{1+e^{-1/\sigma}} e^{-\frac{\sigma^2}{2t^2}} \sum_{y \in \mathbb{Z}} e^{-\frac{y^2}{2\sigma^2}}$.

Proof.

$$\begin{aligned}
E[C] &= E[E[C|Y]] \\
&= E[e^{-\frac{(|Y|-\sigma^2/t)^2}{2\sigma^2}}] && \text{since } E[Bernoulli(p)] = p \\
&= \frac{1-e^{-1/\sigma}}{1+e^{-1/\sigma}} \sum_{y \in \mathbb{Z}} e^{-\frac{(|y|-\sigma^2/t)^2}{2\sigma^2} - |y|/t} && \text{expectation over } Y \sim \mathcal{L}_{\mathcal{Z}}(0, \sigma) \\
&= \frac{1-e^{-1/\sigma}}{1+e^{-1/\sigma}} e^{-\frac{\sigma^2}{2t^2}} \sum_{y \in \mathbb{Z}} e^{-\frac{y^2}{2\sigma^2}}
\end{aligned}$$

□

We now show that conditioning Y on the success of C gives the desired output distribution.

Theorem 2.4. [CKS20] If y is a realization of $Y \sim \mathcal{L}_{\mathbb{Z}}(0, t)$ and c is a realization of $C \sim Bernoulli(exp(-(|y| - \sigma^2/t)^2/(2\sigma^2)))$, then $P[Y = y | C = \top] = \frac{e^{-\frac{y^2}{2\sigma^2}}}{\sum_{y' \in \mathbb{Z}} e^{-\frac{y'^2}{2\sigma^2}}}$. That is, $Y|_{C=\top} \sim \mathcal{N}_{\mathbb{Z}}(0, \sigma^2)$.

Proof.

$$\begin{aligned}
P[Y = y | C = \top] &= \frac{P[C = \top | Y = y] P[Y = y]}{P[C = \top]} && \text{Bayes' Theorem} \\
&= \frac{e^{-\frac{(|y|-\sigma^2/t)^2}{2\sigma^2}} \frac{1-e^{-1/t}}{1+e^{-1/t}} e^{-|y|/t}}{E[C]} && \text{by definition of } \mathcal{L}_{\mathbb{Z}}(0, \sigma) \\
&= \frac{e^{-\frac{(|y|-\sigma^2/t)^2}{2\sigma^2}} e^{-|y|/t}}{e^{-(\sigma/t)^2/2} \sum_{y' \in \mathbb{Z}} e^{-\frac{y'^2}{2\sigma^2}}} && \text{by 2.3} \\
&= \frac{e^{-\frac{y^2}{2\sigma^2}}}{\sum_{y' \in \mathbb{Z}} e^{-\frac{y'^2}{2\sigma^2}}}
\end{aligned}$$

□

Lemma 2.5. If the outcome of `sample_discrete_gaussian` is `Ok(out)`, then `out` is distributed as $\mathcal{N}_{\mathbb{Z}}(0, scale^2)$.

Proof. In the 2.2 proof, it was established that the preconditions on `sample_discrete_laplace` are met, so `candidate` on line 9 is distributed as $\mathcal{L}_{\mathbb{Z}}(0, t)$. Similarly, by the definition of `sample_bernoulli_exp`, the outcome is distributed according to $Bernoulli(exp(-(|y| - \sigma^2/t)^2/(2\sigma^2)))$. Since on line 12, we condition returning `candidate` on a $\top$ sample, the conditions to apply 2.4 are met. Therefore `out` is distributed as $\mathcal{N}_{\mathbb{Z}}(0, scale^2)$. $\square$

Proof. 1 holds by 2.2 and 2.5. $\square$

References

[CKS20] Clément L. Canonne, Gautam Kamath, and Thomas Steinke. The discrete gaussian for differential privacy. *CoRR*, abs/2004.00010, 2020.