

fn sample_geometric_exp_fast

Michael Shoemate

September 25, 2024

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

Proves soundness of `fn sample_geometric_exp_fast` in `mod.rs` at commit `0be3ab3e6` (outdated¹). This proof is an adaptation of subsection 5.2 of [CKS20].

Vetting history

- Pull Request #519

1 Hoare Triple

Precondition

$x \in \mathbb{Q} \wedge x > 0$

Pseudocode

```
1 def sample_geometric_exp_fast(x) -> int:
2     if x == 0:
3         return 0
4
5     s, t = Rational.into_numer_denom(x)
6
7     while True:
8         u = Integer.sample_uniform_int_below(t) #
9         d = bool.sample_bernoulli_exp(Rational(u, t)) #
10        if d:
11            break
12
13        v = sample_geometric_exp_slow(1) #
14        z = u + t * v
15        return z // s
```

Postcondition

For any setting of the input parameter x such that the given preconditions hold, `sample_geometric_exp_fast` either returns `Err(e)` due to a lack of system entropy, or `Ok(out)`, where `out` is distributed as *Geometric*($1 - \exp(-x)$).

¹See new changes with `git diff 0be3ab3e6..924a40fe rust/src/traits/samplers/cks20/mod.rs`

2 Proof

Assume the preconditions are met.

Lemma 2.1. `sample_geometric_exp_fast` only returns `Err(e)` when there is a lack of system entropy.

Proof. `x` is of type `Rational`, there exists some non-negative integer s and positive integer t such that $x = s/t$. This is why `Rational.into_numer_denom` is infallible. Since t is a positive integer, the preconditions on `SampleUniformIntBelow` are met, `sample_uniform_int_below` can only return an error due to lack of system entropy, and u is a non-negative integer. Similarly, the preconditions on `sample_bernoulli_exp` and `sample_geometric_exp_slow` are met, and their definitions guarantee an error is only returned due to lack of system entropy. The only source of errors is from the invocation of these functions, therefore `sample_geometric_exp_fast` only returns `Err(e)` when there is a lack of system entropy. \square

We now establish some lemmas that will be useful in proving the distribution of `out`.

- Let u be a realization of a random variable $U \sim \text{Uniform}(0, t)$, supported on $[0, t)$.
- Let d be a realization of a random variable $D \sim \text{Bernoulli}(\exp(-u/t))$
- Let v be a realization of a random variable $V \sim \text{Geometric}(1 - \exp(-1))$

Lemma 2.2. [CKS20] Conditioned on $d = \top$, if $z = u + t \cdot v$, then z is a realization of a random variable $Z \sim \text{Geometric}(1 - \exp(-1/t))$. Equivalently, $P[Z = z | D = \top] = (1 - e^{-1/t})e^{-z/t}$.

Proof. For any z , define $u_z := z \bmod t$ and $v_z := \lfloor z/t \rfloor$, so that $z = u_z + t \cdot v_z$.

$$\begin{aligned}
 P[Z = z | D = \top] &= P[U = u_z, V = v_z | D = \top] && \text{since } z = u_z + t \cdot v_z \\
 &= P[U = u_z | D = \top] P[V = v_z] && \text{as } U \text{ and } V \text{ are independent} \\
 &= \frac{P[U = u_z]}{P[D = \top]} P[D = \top | U = u_z] \cdot (1 - e^{-1})e^{-v_z} && \text{by Bayes' theorem} \\
 &= \frac{1/t}{1/t \sum_{k=0}^{t-1} e^{-k/t}} e^{-u_z/t} \cdot (1 - e^{-1})e^{-v_z} && \text{since } P[D = \top] = \frac{1}{t} \sum_{k=0}^{t-1} e^{-k/t} \\
 &= \frac{(1 - e^{-1})}{\sum_{k=0}^{t-1} e^{-k/t}} e^{-(u_z/t + v_z)} \\
 &= (1 - e^{-1/t})e^{-(u_z/t + v_z)} \\
 &= (1 - e^{-1/t})e^{-z/t} && \text{since } z = u_z + t \cdot v_z
 \end{aligned}$$

\square

Lemma 2.3. [CKS20] Fix $p \in (0, 1]$. Let G be a $\text{Geometric}(1 - p)$ random variable, and $n \geq 1$ be an integer. Then $\lfloor G/n \rfloor$ is a $\text{Geometric}(1 - q)$ random variable with $q = p^n$.

Proof.

$$\begin{aligned}
 P[\lfloor G/n \rfloor = k] &= P[nk < G < (k+1)n] && \text{any } G \text{ in the interval maps to } k \\
 &= \sum_{l=nk}^{(k+1)n-1} (1-p)p^l \\
 &= (1-p^n)p^{nk} \\
 &= (1-q)q^k
 \end{aligned}$$

\square

Theorem 2.4. [CKS20] Given any $s, t \in \mathbb{Z}_+$ and $Z \sim \text{Geometric}(1 - \exp(-1/t))$, define $Y = \lfloor Z/s \rfloor$. Then $Y \sim \text{Geometric}(1 - \exp(-s/t))$.

Proof.

$$\begin{aligned}
P[Y = y | D = \top] &= P[\lfloor Z/s \rfloor = y | D = \top] \\
&= (1 - p^s) p^{sk} && \text{by 2.3} \\
&= (1 - (e^{-1/t})^s) (e^{-1/t})^{sk} \\
&= (1 - e^{-s/t}) (e^{-s/t})^k
\end{aligned}$$

□

Lemma 2.5. If the outcome of `sample_geometric_exp_fast` is `Ok(out)`, then `out` is distributed as $\text{Geometric}(1 - \exp(-x))$.

Proof. As shown in 2.1, the preconditions for `SampleUniformIntBelow` on line 8, `sample_bernoulli_exp` on line 9, and `sample_bernoulli_exp_slow` on line 13 are met. Therefore, `u`, `d` and `v` follow the distributions necessary to apply 2.2. By 2.2, `z` is a realization of $Z \sim \text{Geometric}(1 - \exp(-1/t))$. Since `z` is a realization of $Z \sim \text{Geometric}(1 - \exp(-1/t))$, then by 2.4, `out` is distributed as $\text{Geometric}(1 - \exp(-x))$. □

Proof. 1 holds by 2.1 and 2.5. □

References

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