1 FastLSH for l_1 distance

For given vector pair $(\boldsymbol{v}, \boldsymbol{u})$, let $s = \|\boldsymbol{v} - \boldsymbol{u}\|_1$. For our purpose, the collection of n entries $|v_i - u_i|$ $\{i = 1, 2, \ldots, n\}$ is viewed as a population, which follows an unknown distribution with finite mean $\mu = (\sum_{i=1}^{n} |v_i - u_i|)/n$ and variance $\sigma^2 = (\sum_{i=1}^{n}(|v_i - u_i| - \mu)^2)/n$. After performing the sampling operator $S(\cdot)$ of size m, $\boldsymbol{v}$ and $\boldsymbol{u}$ are transformed into $\tilde{\boldsymbol{v}} = S(\boldsymbol{v})$ and $\tilde{\boldsymbol{u}} = S(\boldsymbol{u})$, and the l_1 distance of $(\tilde{\boldsymbol{v}}, \tilde{\boldsymbol{u}})$ is $\tilde{s} = \sum_{i=1}^{m} |\tilde{v}_i - \tilde{u}_i|$. By Central Limit Theorem, we have the following lemma:

Lemma 1.1. *If m is sufficiently large, then the sum $\tilde{s}$ of m i.i.d random samples $|\tilde{v}_i - \tilde{u}_i|$ $(i \in 1, 2, \ldots, m)$ converges asymptotically to the normal distribution with mean $m\mu$ and variance $m\sigma^2$, i.e., $\tilde{s} \sim \mathcal{N}(m\mu, m\sigma^2)$.*

Since $\tilde{s} \geq 0$, $\tilde{s}$ can be modeled by normal distribution $\tilde{s} \sim \mathcal{N}(m\mu, m\sigma^2)$ over the truncation interval $[0, +\infty)$, that is, the singly-truncated normal distribution $\psi(x; \tilde{\mu}, \tilde{\sigma}^2, 0, +\infty)$, where $\tilde{\mu} = m\mu$ and $\tilde{\sigma}^2 = m\sigma^2$.

If $\boldsymbol{a}$ is a projection vector with entries being $i.i.d$ samples drawn from standard Cauchy distribution. It follows from the p-stability that the distance between projections $(\boldsymbol{a}^T\boldsymbol{v} - \boldsymbol{a}^T\boldsymbol{u})$ for two vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ is distributed as $\|\boldsymbol{v} - \boldsymbol{u}\|_1 X$, i.e., sX, where X follows standard Cauchy distribution. Similarly, the projection distance between $\tilde{\boldsymbol{v}}$ and $\tilde{\boldsymbol{u}}$ under $\tilde{\boldsymbol{a}}$ $(\tilde{\boldsymbol{a}}^T\tilde{\boldsymbol{v}} - \tilde{\boldsymbol{a}}^T\tilde{\boldsymbol{u}})$ follows the distribution $\tilde{s}X$. Next, we will use the following Lemma 1.2 to determine the distribution of $\tilde{s}X$.

Lemma 1.2. *The characteristic function of the product of two independent random variables $W = XY$ is as follows:*

$$\varphi_W(x) = E_Y\{exp(-|xY|)\}$$

where X is a standard Cauchy random variable and Y is an independent random variable with mean μ and variance σ^2.

Proof. This proof is similar to Lemma A.1 in Appendix. □

Since the distribution of a random variable is determined uniquely by its characteristic function. With respect to $\tilde{s}X$, the characteristic function can be obtained by Lemma 1.3 since X follows Cauchy distribution:

Lemma 1.3. *The characteristic function of $\tilde{s}X$ is as follows:*

$$\begin{aligned}\varphi_{\tilde{s}X}(x) = &\frac{1}{2(1 - \Phi(\frac{-\tilde{\mu}}{\tilde{\sigma}}))} \cdot exp(\frac{-\tilde{\mu}^2 + (\tilde{\mu} - \tilde{\sigma}^2|x|)^2}{2\tilde{\sigma}^2}) \\ &\cdot \operatorname{erfc}(\frac{\tilde{\sigma}^2|x| - \tilde{\mu}}{\sqrt{2}\tilde{\sigma}}) \quad (-\infty < x < +\infty)\end{aligned}$$

where $\operatorname{erfc}(t) = \frac{2}{\sqrt{\pi}}\int_t^{+\infty} \exp(-x^2)dx$ $(-\infty < t < +\infty)$ *is the complementary error function.*

Proof. According to Lemma 1.1 and Lemma 1.2, we have:

$$\begin{aligned}
\varphi_{\tilde{s}X}(x) &= \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \int_0^{+\infty} \frac{\exp(-|x|\,y - \frac{(y-\tilde{\mu})^2}{2\tilde{\sigma}^2})}{\Phi(a_2;\tilde{\mu},\tilde{\sigma}^2) - \Phi(a_1;\tilde{\mu},\tilde{\sigma}^2)} dy \\
&= \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \int_0^{+\infty} \frac{\exp(-\frac{(y^2-2\tilde{\mu}y+2|x|\tilde{\sigma}^2 y+\tilde{\mu}^2)}{2\tilde{\sigma}^2})}{\Phi(a_2;\tilde{\mu},\tilde{\sigma}^2) - \Phi(a_1;\tilde{\mu},\tilde{\sigma}^2)} dy \\
&= \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \int_0^{+\infty} \frac{\exp(\frac{-\tilde{\mu}+(\tilde{\mu}-|x|\tilde{\sigma}^2)^2}{2\sigma^2}) \cdot \exp(-\frac{(y-(\tilde{\mu}-|x|\tilde{\sigma}^2))^2}{2\tilde{\sigma}^2})}{\Phi(a_2;\tilde{\mu},\tilde{\sigma}^2) - \Phi(a_1;\tilde{\mu},\tilde{\sigma}^2)} dy \\
&= \frac{1}{2(1-\Phi(\frac{-\tilde{\mu}}{\tilde{\sigma}}))} \cdot \exp(\frac{-\tilde{\mu}+(\tilde{\mu}-|x|\,\tilde{\sigma}^2)^2}{2\sigma^2}) \cdot \mathrm{erfc}(\frac{|x|\,\tilde{\sigma}^2-\tilde{\mu}}{\sqrt{2}\tilde{\sigma}})
\end{aligned}$$

where $a_2 = \infty$ and $a_1 = 0$. Hence we prove this Lemma □

2 FastLSH for $l_{\frac{1}{2}}$ distance

For given vector pair $(\boldsymbol{v}, \boldsymbol{u})$, let $s = \|\boldsymbol{v} - \boldsymbol{u}\|_{\frac{1}{2}}$. For our purpose, the collection of n entries $(v_i - u_i)^{\frac{1}{2}}$ $\{i = 1, 2, \ldots, n\}$ is viewed as a population, which follows an unknown distribution with finite mean $\mu' = (\sum_{i=1}^{n}(v_i-u_i)^{\frac{1}{2}})/n$ and variance $\sigma^2 = (\sum_{i=1}^{n}((v_i-u_i)^{\frac{1}{2}}-\mu)^2)/n$. After performing the sampling operator $S(\cdot)$ of size m, $\boldsymbol{v}$ and $\boldsymbol{u}$ are transformed into $\tilde{\boldsymbol{v}} = S(\boldsymbol{v})$ and $\tilde{\boldsymbol{u}} = S(\boldsymbol{u})$, and the squared root distance of $(\tilde{\boldsymbol{v}}, \tilde{\boldsymbol{u}})$ is $\tilde{s}^{\frac{1}{2}} = \sum_{i=1}^{m}(\tilde{v}_i - \tilde{u}_i)^{\frac{1}{2}}$. By Central Limit Theorem, we have the following lemma:

Lemma 2.1. *If m is sufficiently large, then the sum $\tilde{s}^{\frac{1}{2}}$ of m $i.i.d$ random samples $(\tilde{v}_i - \tilde{u}_i)^{\frac{1}{2}}$ $(i \in 1, 2, \ldots, m)$ converges asymptotically to the normal distribution with mean $m\mu$ and variance $m\sigma^2$, i.e., $\tilde{s}^{\frac{1}{2}} \sim \mathcal{N}(m\mu, m\sigma^2)$.*

Since $\tilde{s}^{\frac{1}{2}} \geq 0$, $\tilde{s}^{\frac{1}{2}}$ can be modeled by normal distribution $\tilde{s}^{\frac{1}{2}} \sim \mathcal{N}(m\mu, m\sigma^2)$ over the truncation interval $[0, +\infty)$, that is, the singly-truncated normal distribution $\psi(x; \tilde{\mu}, \tilde{\sigma}^2, 0, +\infty)$. Considering the fact that $\tilde{s} \geq 0$, we have $Pr[\tilde{s} < t] = Pr[\tilde{s}^{\frac{1}{2}} < t^{\frac{1}{2}}]$ for any $t > 0$. Therefore, the CDF of $\tilde{s}$, denoted by $F_{\tilde{s}}$, can be computed as follows:

$$\begin{aligned}
F_{\tilde{s}}(t) &= Pr[\tilde{s} < t] = Pr[\tilde{s}^{\frac{1}{2}} < t^{\frac{1}{2}}] \\
&= \int_0^{t^{\frac{1}{2}}} \psi(x; \tilde{\mu}, \tilde{\sigma}^2, 0, \infty) dx
\end{aligned} \tag{1}$$

where $\tilde{\mu} = m\mu$ and $\tilde{\sigma}^2 = m\sigma^2$. Due to the fact that the PDF is the derivative of the CDF, the PDF of $\tilde{s}$, denoted by $f_{\tilde{s}}$, is derived as follows:

$$f_{\tilde{s}}(t) = \frac{d}{dt}[F_{\tilde{s}}(t)] = \frac{1}{2} t^{-\frac{1}{2}} \psi(t^{\frac{1}{2}}; \tilde{\mu}, \tilde{\sigma}^2, 0, \infty) \tag{2}$$

If $\boldsymbol{a}$ is a projection vector with entries being $i.i.d$ samples drawn from standard Lévy distribution. It follows from the p-stability that the distance between projections ($\boldsymbol{a}^T\boldsymbol{v} - \boldsymbol{a}^T\boldsymbol{u}$) for two vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ is distributed as $\|\boldsymbol{v} - \boldsymbol{u}\|_1 X$, i.e., sX, where X follows standard Lévy distribution. Similarly, the projection distance between $\tilde{\boldsymbol{v}}$ and $\tilde{\boldsymbol{u}}$ under $\tilde{\boldsymbol{a}}$ ($\tilde{\boldsymbol{a}}^T\tilde{\boldsymbol{v}} - \tilde{\boldsymbol{a}}^T\tilde{\boldsymbol{u}}$) follows the distribution $\tilde{s}X$. Next, we will use the following Lemma 2.2 to determine the distribution of $\tilde{s}X$.

Lemma 2.2. *The characteristic function of the product of two independent random variables $W = XY$ is as follows:*

$$\varphi_W(x) = E_Y\{exp(-\sqrt{2ixY})\}$$

where X is a standard Lévy random variable and Y is an independent random variable with mean μ and variance σ^2.

Proof. This proof is similar to Lemma A.1 in Appendix. $\square$

Since the distribution of a random variable is determined uniquely by its characteristic function. With respect to $\tilde{s}X$, the characteristic function can be obtained by Lemma 1.3 since X follows Lévy distribution:

Lemma 2.3. *The characteristic function of $\tilde{s}X$ is as follows:*

$$\begin{aligned}\varphi_{\tilde{s}X}(x) = & \frac{1}{2(1-\Phi(\frac{-\tilde{\mu}}{\tilde{\sigma}}))} \cdot exp(\frac{-\tilde{\mu}^2 + (\tilde{\mu} - \sqrt{-2ix}\tilde{\sigma}^2)^2}{2\tilde{\sigma}^2}) \\ & \cdot \operatorname{erfc}(\frac{\sqrt{-2ix}\tilde{\sigma}^2 - \tilde{\mu}}{\sqrt{2}\tilde{\sigma}}) \quad (-\infty < x < +\infty)\end{aligned}$$

where $\operatorname{erfc}(t) = \frac{2}{\sqrt{\pi}}\int_t^{+\infty} \exp(-x^2)dx$ $(-\infty < t < +\infty)$ *is the complementary error function.*

Proof. According to Eqn. (2) and Lemma 2.2, we have:

$$\begin{aligned}\varphi_{\tilde{s}X}(x) &= \frac{1}{\sqrt{2\pi}\tilde{\sigma}}\int_0^{+\infty} \frac{\exp(-\sqrt{-2ixy} - \frac{(\sqrt{y}-\tilde{\mu})^2}{2\tilde{\sigma}^2})}{2\sqrt{y}(\Phi(a_2;\tilde{\mu},\tilde{\sigma}^2) - \Phi(a_1;\tilde{\mu},\tilde{\sigma}^2))}dy \\ &= \frac{1}{\sqrt{2\pi}\tilde{\sigma}}\int_0^{+\infty} \frac{\exp(-\sqrt{-2ixy^2} - \frac{(y-\tilde{\mu})^2}{2\tilde{\sigma}^2})}{\Phi(a_2;\tilde{\mu},\tilde{\sigma}^2) - \Phi(a_1;\tilde{\mu},\tilde{\sigma}^2)}dy \\ &= \frac{1}{\sqrt{2\pi}\tilde{\sigma}}\int_0^{+\infty} \frac{\exp(-\frac{(y^2-2\tilde{\mu}y+2\sqrt{-2ix}\tilde{\sigma}^2y+\tilde{\mu}^2)}{2\tilde{\sigma}^2})}{\Phi(a_2;\tilde{\mu},\tilde{\sigma}^2) - \Phi(a_1;\tilde{\mu},\tilde{\sigma}^2)}dy \\ &= \frac{1}{\sqrt{2\pi}\tilde{\sigma}}\int_0^{+\infty} \frac{\exp(\frac{-\tilde{\mu}+(\tilde{\mu}-\sqrt{-2ix}\tilde{\sigma}^2)^2}{2\sigma^2}) \cdot \exp(-\frac{(y-(\tilde{\mu}-\sqrt{-2ix}\tilde{\sigma}^2))^2}{2\tilde{\sigma}^2})}{\Phi(a_2;\tilde{\mu},\tilde{\sigma}^2) - \Phi(a_1;\tilde{\mu},\tilde{\sigma}^2)}dy \\ &= \frac{1}{2(1-\Phi(\frac{-\tilde{\mu}}{\tilde{\sigma}}))} \cdot \exp(\frac{-\tilde{\mu} + (\tilde{\mu} - \sqrt{-2ix}\tilde{\sigma}^2)^2}{2\sigma^2}) \cdot \operatorname{erfc}(\frac{\sqrt{-2ix}\tilde{\sigma}^2 - \tilde{\mu}}{\sqrt{2}\tilde{\sigma}})\end{aligned}$$

where $a_2 = \infty$ and $a_1 = 0$. Hence we prove this Lemma $\square$