Supplementary Material
A Mathematical Analysis of Learning Loss for Active Learning in Regression

Megh Shukla
Mercedes-Benz Research and Development India
megh.shukla@daimler.com

Abstract

This supplementary material contains the derivations that support the content of the main paper. We first derive a known result that under certain conditions the integral of the gamma distribution has a closed form solution. This solution is useful in computing the probability of sampling a pair of values within δ (Eq: 5, main paper). We then derive the gradient (Eq: 4, main paper) and its expectation (Eq: 6, main paper) of the LearningLoss++ objective.

1. Integral of the Gamma Distribution

Our goal is to compute the integral of the gamma distribution:

$$\int \gamma(x, k, \Theta)\mathrm{d}x = -\Theta \sum_{n=1}^{k} \frac{x^{n-1}e^{-\frac{x}{\Theta}}}{\Theta^n \Gamma(n)} = G(x, k, \Theta) \tag{1}$$

Although this is a known result, we provide a brief outline of the proof. The assumption is that k, the shape parameter $\in \mathbb{Z}^+$. We need to simplify the following:

$$\int \gamma(x, k, \Theta)\mathrm{d}x = \int \frac{x^{k-1}e^{-\frac{x}{\Theta}}}{\Theta^k \Gamma(k)}\mathrm{d}x$$

Using integration by parts,

$$\int \gamma(x, k, \Theta)\mathrm{d}x = -\Theta \frac{x^{k-1}e^{-\frac{x}{\Theta}}}{\Theta^k \Gamma(k)} + \int \frac{x^{k-2}e^{-\frac{x}{\Theta}}}{\Theta^{k-1} \Gamma(k-1)}\mathrm{d}x$$

The above equation can be written as $\int \gamma(x, k, \Theta)\mathrm{d}x = -\Theta\gamma(x, k, \Theta) + \int \gamma(x, k-1, \Theta)\mathrm{d}x$. By recursively solving the integral term using integration by parts, the equation reduces to $\int \gamma(x, k, \Theta)\mathrm{d}x = -\Theta\gamma(x, k, \Theta) - \Theta\gamma(x, k-1, \Theta) \ldots - \Theta\gamma(x, k=1, \Theta)$. Hence, we can write the final form of the equation as:

$$\int \gamma(x, k, \Theta)\mathrm{d}x = -\Theta \sum_{n=1}^{k} \frac{x^{n-1}e^{-\frac{x}{\Theta}}}{\Theta^n \Gamma(n)} \tag{2}$$

2. Closed form solution for $P(|X-Y| \leq \delta)$ when $X, Y \sim \gamma(k, \Theta)$, $k \in \mathbb{Z}^+$

The probability of sampling two variables within a margin δ of each other can be written as:

$$P(|X-Y| \leq \delta) = \int_0^{\delta} \gamma(x,k,\Theta) \int_0^{x+\delta} \gamma(y,k,\Theta)\mathrm{d}y\mathrm{d}x + \int_{\delta}^{\infty} \gamma(x,k,\Theta) \int_{x-\delta}^{x+\delta} \gamma(y,k,\Theta)\mathrm{d}y\mathrm{d}x \tag{3}$$

Using the previous result Eq: 2, we can simplify the above equation as:

$$= \int_0^{\delta} \gamma(x,k,\Theta) \left[-\Theta \sum_{n=1}^{k} \frac{(x+\delta)^{n-1} e^{-\frac{(x+\delta)}{\Theta}}}{\Theta^n \Gamma(n)} \right] \mathrm{d}x + \int_0^{\delta} \gamma(x,k,\Theta)\mathrm{d}x$$
$$+ \int_{\delta}^{\infty} \gamma(x,k,\Theta) \left[-\Theta \sum_{n=1}^{k} \frac{(x+\delta)^{n-1} e^{-\frac{(x+\delta)}{\Theta}}}{\Theta^n \Gamma(n)} \right] \mathrm{d}x - \int_{\delta}^{\infty} \gamma(x,k,\Theta) \left[-\Theta \sum_{n=1}^{k} \frac{(x-\delta)^{n-1} e^{-\frac{(x-\delta)}{\Theta}}}{\Theta^n \Gamma(n)} \right] \mathrm{d}x$$

We write Eq: 4 as $P(|X-Y| \leq \delta) = A+B+C+D$. While a closed form solution can be easily obtained for B (integral of gamma), We rely on the use of the binomial theorem $(x+\delta)^{n-1} = \sum_{i=0}^{i=n-1} {}^{n-1}C_i x^i \delta^{(n-1)-i}$ to simplify terms A, C and D. The method to solve A, C and D remains the same, hence we show here to focus on reducing A to a closed form solution in this supplementary material.

$$A = \int_0^{\delta} \gamma(x,k,\Theta) \left[-\Theta \sum_{n=1}^{k} \frac{(x+\delta)^{n-1} e^{-\frac{(x+\delta)}{\Theta}}}{\Theta^n \Gamma(n)} \right] \mathrm{d}x$$
$$= -\Theta \sum_{n=1}^{k} \sum_{i=0}^{n-1} \int_{x=0}^{\delta} \frac{{}^{n-1}C_i \delta^{(n-1)-i} x^{k+i-1} e^{-\frac{2x+\delta}{\Theta}}}{\Theta^{k+n}\Gamma(n)\Gamma(k)} \mathrm{d}x$$

We try to write the above equations by creating a new gamma distribution. After reorganizing the terms, we get:

$$A = -\Theta e^{-\dfrac{\delta}{\Theta}} \sum_{n=1}^{k} \sum_{i=0}^{n-1} \frac{{}^{n-1}C_i \delta^{(n-1)-i}}{2^{k+i}\Theta^{n-i}} \int_{x=0}^{\delta} \frac{x^{k+i-1} e^{-\frac{x}{\Theta/2}}}{(\dfrac{\Theta}{2})^{k+i}\Gamma(n)\Gamma(k)} \mathrm{d}x$$

The final obstacle of writing the integral term into another gamma distribution is introducing $\Gamma(k+i)$. We use the property of gamma functions, $\Gamma(k+i) = \dfrac{(k+i-1)!}{(k-1)!}\Gamma(k)$. We reduce A to:

$$A = -\Theta e^{-\dfrac{\delta}{\Theta}} \sum_{n=1}^{k} \frac{1}{\Gamma(n)} \sum_{i=0}^{n-1} \frac{{}^{n-1}C_i \; {}^{k+i-1}P_i \delta^{(n-1)-i}}{2^{k+i}\Theta^{n-i}} \int_{x=0}^{\delta} \gamma(x, k+i, \Theta/2)\mathrm{d}x$$

Fortunately, we have previously shown that a closed form solution exists to compute the integral of the gamma function. We therefore reach the final solution for A:

$$A = -\Theta e^{-\dfrac{\delta}{\Theta}} \sum_{n=1}^{k} \frac{1}{\Gamma(n)} \sum_{i=0}^{n-1} \frac{{}^{n-1}C_i \; {}^{k+i-1}P_i \delta^{(n-1)-i}}{2^{k+i}\Theta^{n-i}} (-\frac{\Theta}{2} \sum_{m=1}^{k+i} \frac{\delta^{m-1} e^{-\frac{\delta}{\Theta/2}}}{(\Theta/2)^m \Gamma(m)} + 1)$$

B, C, D involve a similar reduction process. Evaluation of terms at $lim\ x \to \infty$ reduces to 0 since all the terms contain (x^n/e^x). We reproduce the final solution for A, B, C, D below:

$$
\begin{aligned}
A &= -\Theta e^{-\dfrac{\delta}{\Theta}} \sum_{n=1}^{k} \frac{1}{\Gamma(n)} \sum_{i=0}^{n-1} \frac{{}^{n-1}C_i \; {}^{k+i-1}P_i \delta^{(n-1)-i}}{2^{k+i}\Theta^{n-i}} \Big(-\frac{\Theta}{2} \sum_{m=1}^{k+i} \frac{\delta^{m-1} e^{-\frac{\delta}{\Theta/2}}}{(\Theta/2)^m \Gamma(m)} + 1\Big) \\
B &= -\Theta \sum_{n=1}^{k} \frac{\delta^{n-1} e^{-\frac{\delta}{\Theta}}}{\Theta^n \Gamma(n)} + 1 \\
C &= -\Theta e^{-\dfrac{\delta}{\Theta}} \sum_{n=1}^{k} \frac{1}{\Gamma(n)} \sum_{i=0}^{n-1} \frac{{}^{n-1}C_i \; {}^{k+i-1}P_i \delta^{(n-1)-i}}{2^{k+i}\Theta^{n-i}} \Big(\frac{\Theta}{2} \sum_{m=1}^{k+i} \frac{\delta^{m-1} e^{-\frac{\delta}{\Theta/2}}}{(\Theta/2)^m \Gamma(m)}\Big) \\
D &= \Theta e^{\dfrac{\delta}{\Theta}} \sum_{n=1}^{k} \frac{1}{\Gamma(n)} \sum_{i=0}^{n-1} \frac{{}^{n-1}C_i \; {}^{k+i-1}P_i (-\delta)^{(n-1)-i}}{2^{k+i}\Theta^{n-i}} \Big(\frac{\Theta}{2} \sum_{m=1}^{k+i} \frac{\delta^{m-1} e^{-\frac{\delta}{\Theta/2}}}{(\Theta/2)^m \Gamma(m)}\Big)
\end{aligned} \tag{4}
$$

As in the paper, let $f(i,n,k,\delta,\Theta) = \frac{e^{-\frac{\delta}{\Theta}} \; {}^{n-1}C_i \; {}^{k+i-1}P_i \delta^{(n-1)-i}}{2^{k+i}\Theta^{n-i}}$ and $G(x,k,\Theta) = -\Theta \sum_{n=1}^{k} \frac{x^{n-1} e^{-\frac{x}{\Theta}}}{\Theta^n \Gamma(n)}$. We note the C cancels the first term in A, leaving us with the final solution for $P(|X-Y| \leq \delta)$ = A + B + C + D:

$$
\begin{aligned}
P(|X-Y| \leq \delta) = 1 - \Theta G(\delta,k,\Theta) - \Theta \sum_{n=1}^{n=k} \frac{1}{\Gamma(n)} \sum_{i=0}^{i=n-1} f(i,n,k,\delta,\Theta) \\
+ \Theta \sum_{n=1}^{n=k} \frac{1}{\Gamma(n)} \sum_{i=0}^{i=n-1} f(i,n,k,-\delta,\Theta) G(\delta,k+i,\Theta/2)
\end{aligned} \tag{5}
$$

3. LearningLoss++ Gradient

We define similar notations from the paper: (l_i, l_j) represent the true loss for images (x_i, x_j), the intermediate representations from the network for these images being (θ_i, θ_j). We define the learning loss network to be $\hat{l}_i = \theta_i^T w$ where $\hat{l}_i$ is the predicted/indicative loss for image x_i. We define the ground truth probability of sampling x_i over x_j as: $p_i = l_i/(l_i + l_j)$ and similarly for p_j. The network's probability of sampling x_i over x_j is $q_i = e^{\hat{l}_i}/(e^{\hat{l}_i} + e^{\hat{l}_j})$ with q_j defined similarly. The minimizatio objective is:

$$\mathbb{L}_{loss}(w, \theta_i, \theta_j) = \text{KL}(p||q) = p_i \text{log}\frac{p_i}{q_i} + p_j \text{log}\frac{p_j}{q_j} \tag{6}$$

On substituting p, q and computing the gradient with respect to w, Eq: 6 reduces to:

$$\begin{aligned}
\nabla_w \mathbb{L} &= -\nabla_w \left[p_i log(\frac{e^{\theta_i^T w}}{e^{\theta_i^T w} + e^{\theta_j^T w}}) + p_j log(\frac{e^{\theta_j^T w}}{e^{\theta_i^T w} + e^{\theta_j^T w}}) \right] \\
&= -\nabla_w \left[p_i \theta_i^T w + p_j \theta_j^T w - (p_i + p_j) \text{log}(e^{\theta_i^T w} + e^{\theta_j^T w}) \right] \\
&= -p_i \theta_i - p_j \theta_j + \frac{e^{\theta_i^T w} \theta_i + e^{\theta_j^T w} \theta_j}{e^{\theta_i^T w} + e^{\theta_j^T w}}
\end{aligned}$$

Using the definition of $\hat{l}_i, \hat{l}_j, q_i, q_j$, the equation can be written as:

$$\nabla_w \mathbb{L} = -p_i \theta_i - p_j \theta_j + q_i \theta_i + q_j \theta_j \tag{7}$$

Since $p_i + p_j = 1$, $q_i + q_j = 1$, we get $(q_i - p_i) = -(q_j - p_j)$. The final gradient can now be written as:

$$\nabla_w \mathbb{L}(w, \theta_i, \theta_j) = (q_i - p_i)(\theta_i - \theta_j) \tag{8}$$

4. Expected Gradient for LearningLoss++

Since providing a proof for the entire solution is time consuming and lengthy, we provide a derivation for the main skeleton and show that the solutions discussed above (integral of gamma, binomial) can be reused to obtain a closed form solution for the expected gradient. We continue from Eq: 8 in the paper; the expected gradient is defined as:

$$\mathbb{E}_x[\nabla_w \mathbb{L}(X = x, Y = x + \delta_2 \,|\, \delta_2)] = \int_{x=0}^{x=\infty} \int_{y=x+\delta_1}^{y=x+\delta_2} (q_i - \frac{x}{2x + \delta_2})(\theta_i - \theta_j) p(x, y|\delta_2) \mathrm{d}y \mathrm{d}x \tag{9}$$

Where we define δ_1 as $\lim \delta_2 - \delta_1 \to 0^+$ to accurately define area under the curve as probability. By definition, $p(x, y|\delta_2) = \dfrac{\gamma(x, k, \Theta)\gamma(y, k, \Theta)}{p(y - x = \delta_2)}$, since $X, Y \sim \gamma(x, k, \Theta)$. Here, $p(y - x = \delta_2)$ is the normalizer. We note that $p(y - x = \delta_2) = \int_{x=0}^{\infty} \int_{y=x+\delta_1}^{x+\delta_2} \gamma(x, k, \Theta)\gamma(y, k, \Theta) \mathrm{d}y \mathrm{d}x$ and $\delta_1 \to \delta_2^-$. We simplify $p_i = \dfrac{x}{2x + \delta_2} = \dfrac{1}{2}(1 - \dfrac{\delta_2}{2x + \delta_2})$. The expectation reduces to:

$$\mathbb{E}_x[\nabla_w \mathbb{L}] = q_i(\theta_i - \theta_j) - \frac{(\theta_i - \theta_j)}{2} \left[1 - \int_{x=0}^{\infty} \frac{\delta_2}{2x + \delta_2} \int_{y=x+\delta_1}^{x+\delta_2} \frac{\gamma(x, k, \Theta)\gamma(y, k, \Theta)}{p(y - x = \delta_2)} \mathrm{d}y \mathrm{d}x \right] \tag{10}$$

Since $p(y - x = \delta_2)$ is the normalizer, it is constant given δ_2. We therefore write $\mathbb{D} = p(y - x = \delta_2)$. This allows us to write Eq: 10 as:

$$= q_i(\theta_i - \theta_j) - \frac{(\theta_i - \theta_j)}{2} \left[1 - \int_{x=0}^{\infty} \frac{\delta_2}{2x + \delta_2} \frac{\gamma(x, k, \Theta)}{\mathbb{D}} \int_{y=x+\delta_1}^{x+\delta_2} \gamma(y, k, \Theta) \mathrm{d}y \mathrm{d}x \right] \tag{11}$$

We see that Eq: 11 bears a strong resemblance with the derivation of $P(|X - Y| \leq \delta)$ we proved earlier. We can directly substitute the values of $\gamma(x, k, \Theta)$, $\int_{y=x+\delta_1}^{x+\delta_2} \gamma(y, k, \Theta)$ [Integral of Gamma] and $f(i, n, k, \delta, \Theta)$ from Eq: 5 into the above equation to get:

$$\begin{aligned} = q_i(\theta_i - \theta_j) - \frac{(\theta_i - \theta_j)}{2} \; [\; 1 + \frac{\Theta}{\mathbb{D}} \sum_{n=1}^{k} \frac{1}{\Gamma(n)} \sum_{i=0}^{n-1} f(i, n, k, \delta_2, \Theta) \int_{x=0}^{\infty} \frac{x^{k+i-1} e^{-\frac{x}{\Theta/2}}}{(\frac{\Theta}{2})^{k+i}\Gamma(k+i)} \frac{\delta_2}{2x + \delta_2} \mathrm{d}x \\ - \frac{\Theta}{\mathbb{D}} \sum_{n=1}^{k} \frac{1}{\Gamma(n)} \sum_{i=0}^{n-1} f(i, n, k, \delta_1, \Theta) \int_{x=0}^{\infty} \frac{x^{k+i-1} e^{-\frac{x}{\Theta/2}}}{(\frac{\Theta}{2})^{k+i}\Gamma(k+i)} \frac{\delta_2}{2x + \delta_2} \mathrm{d}x \;] \end{aligned}$$

Let $t = 2x + \delta_2$, then the above equation reduces to:

$$\begin{aligned} = q_i(\theta_i - \theta_j) - \frac{(\theta_i - \theta_j)}{2} \; [\; 1 + \frac{\Theta}{\mathbb{D}} \sum_{n=1}^{k} \frac{1}{\Gamma(n)} \sum_{i=0}^{n-1} f(i, n, k, \delta_2, \Theta) \int_{t=\delta_2}^{\infty} \frac{(t - \delta_2)^{k+i-1} e^{-\frac{(t-\delta_2)}{\Theta/2}}}{(\frac{\Theta}{2})^{k+i}\Gamma(k+i)} \frac{\delta_2}{t} \mathrm{d}t \\ - \frac{\Theta}{\mathbb{D}} \sum_{n=1}^{k} \frac{1}{\Gamma(n)} \sum_{i=0}^{n-1} f(i, n, k, \delta_1, \Theta) \int_{t=\delta_2}^{\infty} \frac{(t - \delta_2)^{k+i-1} e^{-\frac{(t-\delta_2)}{\Theta/2}}}{(\frac{\Theta}{2})^{k+i}\Gamma(k+i)} \frac{\delta_2}{t} \mathrm{d}t \;] \end{aligned}$$

If we let $I(k + i, \Theta) = \int_{t=\delta_2}^{\infty} \dfrac{(t - \delta_2)^{k+i-1} e^{-\frac{(t-\delta_2)}{\Theta/2}}}{(\frac{\Theta}{2})^{k+i}\Gamma(k+i)} \dfrac{\delta_2}{t} \mathrm{d}t$, then we can write the expected gradient $\mathbb{E}_x[\nabla_w \mathbb{L}]$ as:

$$\mathbb{E}_x[\nabla_w \mathbb{L}(\delta)] = (\theta_i - \theta_j)\Bigg[q_i - \frac{1}{2} + \frac{\Theta}{2\mathbb{D}} \sum_{n=1}^{k} \frac{1}{\Gamma(n)} \sum_{i=0}^{n-1} I(k+i,\Theta)[f(i,n,k,\delta_2,\Theta) - f(i,n,k,\delta_1,\Theta)]\Bigg] \tag{12}$$

We note that **this is the final expected gradient given the margin** δ and $\delta_1 \rightarrow \delta_2^- = \delta$. However, we still need to compute the closed form solution for $\mathbb{D}$ and $I(k+i,\Theta)$. We first compute the value of $\mathbb{D}$:

$$\mathbb{D} = p(y - x = \delta_2) = \int_{x=0}^{\infty} \gamma(x,k,\Theta) \int_{y=x+\delta_1}^{x+\delta_2} \gamma(y,k,\Theta)\mathrm{d}y\mathrm{d}x$$

We have previously computed a similar result when deriving the closed form solution, where use the integral of gamma as well as the binomial theorem to solve for integrating a gamma function within a gamma function. To avoid repetitive steps, we present the final solution for $\mathbb{D}$:

$$\mathbb{D} = p(y - x = \delta_2) = \Theta \sum_{n=1}^{k} \frac{1}{\Gamma(n)} \sum_{i=0}^{n-1} f(i,n,k,\delta_1,\Theta) - f(i,n,k,\delta_2,\Theta) \tag{13}$$

The solution for $I(k+i,\Theta) = \int_{t=\delta_2}^{\infty} \frac{(t-\delta_2)^{k+i-1} e^{-\frac{(t-\delta_2)}{\Theta/2}}}{(\frac{\Theta}{2})^{k+i}\Gamma(k+i)} \frac{\delta_2}{t}\mathrm{d}t$ is similar with the use of the binomial theorem to convert the integral into a sum of integrals. However, the following caveat exists: The division by t renders one term in the expansion an exponential integral of the form $\frac{e^{-t}}{t}$. This is reflected in the solution for $I(k+i,\Theta)$:

$$I(u = k+i,\Theta) = e^{\frac{\delta_2}{\Theta}} \sum_{j=1}^{u-1} \frac{(-1)^{u-1-j}\delta_2^{u-j}\; {}^{u-1}C_j}{\theta^{u-j}\; {}^{u-1}P_{u-j}} \int_{t=\delta_2}^{\infty} \gamma(j,\Theta) \;+\; \frac{e^{\frac{\delta_2}{\Theta}}(-1)^{u-1}\delta_2^u}{\Theta^u (u-1)!}\Gamma(0,\frac{\delta_2}{\Theta}) \tag{14}$$

While the first term again contains the integral of the gamma function which is a closed form solution, the second term is a consequence of the exponential integral that leads to the lower incomplete gamma function. We therefore have shown that both $\mathbb{D}$ and $I(u = k+i,\Theta)$ have closed form solutions, allowing the expected gradient Eq: 12 to have a closed form solution.