Appendix B. Full proofs for the theorems in the mathematical framework

Theorem 1 (Simplified marginal estimated posterior with uniform prior)**.** *The marginal estimated posterior distribution $q(\theta_i \mid c)$ given class c is proportional to the product of the marginal likelihood $p(c \mid \theta_i)$ and the marginal assumed prior $q(\theta_i)$, i.e.,*

$$q(\theta_i \mid c) \propto p(c \mid \theta_i) q(\theta_i) \text{ for } -3 \leq \theta_i \leq 3$$

Proof.

$$\begin{aligned}
q(\theta_i \mid c) &= \int \cdots \int q(\boldsymbol{\Theta} \mid c)\, d\boldsymbol{\Theta}_{\setminus i} && \text{(marginal estimated posterior, Definition 8)} \\
&= \int \cdots \int \frac{p(c \mid \boldsymbol{\Theta}) q(\boldsymbol{\Theta})}{p(c)}\, d\boldsymbol{\Theta}_{\setminus i} && \text{(estimated posterior using Bayes' theorem, Definition 7)} \\
&= \frac{q(\boldsymbol{\Theta})}{p(c)} \int \cdots \int p(c \mid \boldsymbol{\Theta})\, d\boldsymbol{\Theta}_{\setminus i} && \text{(since } q(\boldsymbol{\Theta}) \text{ is a non-zero constant within } [-3,3], \text{ Definition 2)} \\
&= \frac{q(\boldsymbol{\Theta})}{p(c)} p(c \mid \theta_i) && \text{(marginal likelihood, Definition 6)} \\
&= \frac{q(\theta_i)^d}{p(c)} p(c \mid \theta_i) && \text{(dimensional power scaling of uniform PDF, Definition 2)} \\
&= \frac{q(\theta_i)^{d-1}}{p(c)} p(c \mid \theta_i) q(\theta_i) && \text{(splitting the marginal priors)}
\end{aligned}$$

By considering $\frac{q(\theta_i)^{d-1}}{p(c)}$ as a normalization factor, we obtain:

$$q(\theta_i \mid c) \propto p(c \mid \theta_i) q(\theta_i) \text{ for } -3 \leq \theta_i \leq 3$$

as desired.

□

Theorem 2 (Simplified marginal true posterior under diagonal covariance matrices assumption)**.** *The approximated marginal true posterior distribution $\hat{p}(\theta_i \mid c)$ for each component θ_i, under the assumption of diagonal covariance matrices for both the true prior $p(\boldsymbol{\Theta})$ and the likelihood $p(c \mid \boldsymbol{\Theta})$, is proportional to the product of the marginal true prior $p(\theta_i)$ and the marginal likelihood $p(c \mid \theta_i)$:*

$$\hat{p}(\theta_i \mid c) \propto p(c \mid \theta_i) p(\theta_i).$$

Proof.

$$\hat{p}(\theta_i \mid c) = \int \cdots \int p(\boldsymbol{\Theta} \mid c)\, d\boldsymbol{\Theta}_{\setminus i}$$

(marginal true posterior, Definition 10)

$$= \int \cdots \int \frac{p(c \mid \boldsymbol{\Theta}) p(\boldsymbol{\Theta})}{p(c)}\, d\boldsymbol{\Theta}_{\setminus i}$$

(true posterior using Bayes' theorem)

$$= \frac{1}{p(c)} \int \cdots \int p(c \mid \boldsymbol{\Theta}) p(\boldsymbol{\Theta})\, d\boldsymbol{\Theta}_{\setminus i}$$

(rearranging terms)

$$= \frac{1}{p(c)} \int \cdots \int \prod_{j=1}^{n} p(c \mid \theta_j) \prod_{j=1}^{n} p(\theta_j)\, d\boldsymbol{\Theta}_{\setminus i}$$

(diagonal covariance assumption, Definition 11)

$$= \frac{1}{p(c)} \int \cdots \int \prod_{j=1}^{n} p(c \mid \theta_j) p(\theta_j)\, d\boldsymbol{\Theta}_{\setminus i}$$

(associativity of multiplication)

$$= \frac{1}{p(c)} \int \cdots \int p(c \mid \theta_i) p(\theta_i) \prod_{j \neq i} p(c \mid \theta_j) p(\theta_j)\, d\boldsymbol{\Theta}_{\setminus i}$$

(associativity of multiplication)

$$= \frac{1}{p(c)} p(c \mid \theta_i) p(\theta_i) \int \cdots \int \prod_{j \neq i} p(c \mid \theta_j) p(\theta_j)\, d\boldsymbol{\Theta}_{\setminus i}$$

(since $p(c \mid \theta_i)$ and $p(\theta_i)$ are constant with respect to the marginalization over $\boldsymbol{\Theta}_{\setminus i}$)

$$= \frac{\int \cdots \int \prod_{j \neq i} p(c \mid \theta_j) p(\theta_j)\, d\boldsymbol{\Theta}_{\setminus i}}{p(c)} p(c \mid \theta_i) p(\theta_i)$$

(rearranging terms)

By considering $\frac{\int \cdots \int \prod_{j \neq i} p(c|\theta_j) p(\theta_j)\, d\boldsymbol{\Theta}_{\setminus i}}{p(c)}$ as a normalization factor, we obtain:

$$\hat{p}(\theta_i \mid c) \propto p(c \mid \theta_i) p(\theta_i).$$

Thus, the proportionality relationship is established as desired.

□

Theorem 3 (Marginal estimated posterior as univariate Gaussian)**.** *The marginal estimated posterior $q(\theta_i \mid c)$ for each parameter θ_i is a univariate Gaussian distribution with mean μ_ℓ and covariance σ_ℓ:*

$$q(\theta_i \mid c) = \mathcal{N}(\theta \mid \mu_\ell, {\sigma_\ell}^2) \text{ for } -3 \leq \theta_i \leq 3$$

Proof. From Definition 6, we have that the marginal likelihood $p(c \mid \theta_i)$ follows a Gaussian distribution with mean μ_ℓ and variance σ_ℓ^2, i.e.,

$$p(c \mid \theta_i) = \mathcal{N}(\theta \mid \mu_\ell, {\sigma_\ell}^2)$$

From Theorem 1, we know that, within the parameter plausible range, the marginal posterior distribution $q(\theta_i \mid c)$ given class c and a uniform prior $q(\theta_i)$ is proportional to the product of the marginal likelihood $p(c \mid \theta_i)$ and the marginal prior $q(\theta_i)$, i.e.,

$$q(\theta_i \mid c) \propto p(c \mid \theta_i) q(\theta_i) \text{ for } -3 \leq \theta_i \leq 3$$

Substituting the expression for $p(c \mid \theta_i)$ into the proportionality, we get:

$$q(\theta_i \mid c) \propto \mathcal{N}(\theta \mid \mu_\ell, {\sigma_\ell}^2) \cdot q(\theta_i) \text{ for } -3 \leq \theta_i \leq 3$$

Since $q(\theta_i)$ is a non-zero constant within the region, we can drop the proportionality sign. Therefore,

$$q(\theta_i \mid c) = \mathcal{N}(\theta \mid \mu_\ell, {\sigma_\ell}^2) \text{ for } -3 \leq \theta_i \leq 3$$

Thus, the theorem is proved.

□

Theorem 4 (Approximated marginal true posterior as univariate Gaussian)**.** *The approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ for each parameter θ_i is a univariate Gaussian distribution represented as:*

$$\hat{p}(\theta_i \mid c) = \mathcal{N}(\theta_i \mid \mu_\phi, \sigma_\phi^2)$$

where $\mu_\phi = \frac{\mu_p \sigma_\ell^2 + \mu_\ell \sigma_p^2}{\sigma_\ell^2 + \sigma_p^2}$ is the mean value of the posterior and $\sigma_\phi^2 = \frac{\sigma_p^2 \sigma_\ell^2}{\sigma_p^2 + \sigma_\ell^2}$ is the variance of the posterior.

Proof. From Definition 4, we know that the marginal true prior $p(\theta_i)$ follows a Gaussian distribution with mean μ_p and variance σ_p^2, i.e.,

$$p(\theta_i) = \mathcal{N}(\theta_i \mid \mu_\mathrm{p}, \sigma_\mathrm{p}^2)$$

From Definition 6, we know that the marginal likelihood $p(c \mid \theta_i)$ also follows a Gaussian distribution with mean μ_ℓ and variance σ_ℓ^2, i.e.,

$$p(c \mid \theta_i) = \mathcal{N}(\theta \mid \mu_\ell, \sigma_\ell^2)$$

Now, using Definition 11, the approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ is proportional to the product of the marginal likelihood and the marginal true prior:

$$\hat{p}(\theta_i \mid c) \propto p(c \mid \theta_i) p(\theta_i)$$

Substituting the expressions for $p(c \mid \theta_i)$ and $p(\theta_i)$, we get:

$$\hat{p}(\theta_i \mid c) \propto \mathcal{N}(\theta \mid \mu_\ell, \sigma_\ell^2) \mathcal{N}(\theta_i \mid \mu_\mathrm{p}, \sigma_\mathrm{p}^2)$$

The product of two Gaussian functions results in another Gaussian function, as shown in Choudhary et al. [2021], which provides the general form for this product. Here, the mean μ_ϕ and the variance σ_ϕ^2 of the approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ is given by:

$$\mu_\phi = \frac{\mu_{\mathrm{p}}\sigma_\ell^2 + \mu_\ell\sigma_{\mathrm{p}}^2}{\sigma_\ell^2 + \sigma_{\mathrm{p}}^2}$$

$$\sigma_\phi^2 = \frac{\sigma_{\mathrm{p}}^2\sigma_\ell^2}{\sigma_{\mathrm{p}}^2 + \sigma_\ell^2}$$

Thus, we conclude that the approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ for each parameter θ_i is a univariate Gaussian distribution, represented as:

$$\hat{p}(\theta_i \mid c) = \mathcal{N}(\theta_i \mid \mu_\phi, \sigma_\phi^2)$$

as desired.

□

Lemma 1 (Kullback-Leibler divergence between a normal distribution and a uniform distribution)**.** *The Kullback-Leibler divergence (KLD) between a normal distribution* $x(\theta) = \mathcal{N}(\theta \mid \mu, \sigma^2)$ *and a uniform distribution* $y(\theta) = \mathit{Uniform}(-a, a)$ *over the interval* $[-a, a]$ *is given by:*

$$D_{KL}(x(\theta) \parallel y(\theta)) = \log\left(\frac{1}{\sigma}\right) + \log(2a)\left(\Phi\left(\frac{a-\mu}{\sigma}\right) - \Phi\left(\frac{-a-\mu}{\sigma}\right)\right) - \left(\frac{1}{2} + \log\left(\sqrt{2\pi}\right)\right)$$

where Φ *denotes the cumulative distribution function (CDF) of the standard normal distribution.*

Proof. Starting with the definition of KLD, we have:

$$D_{\mathrm{KL}}(x(\theta) \parallel y(\theta)) = \int_\theta x(\theta) \log\left(\frac{x(\theta)}{y(\theta)}\right) d\theta$$

Substituting the expressions for $x(\theta)$ and $y(\theta)$:

$$D_{\mathrm{KL}}(x(\theta) \parallel y(\theta)) = \int_\theta \mathcal{N}(\theta \mid \mu, \sigma^2) \log\left(\frac{\mathcal{N}(\theta \mid \mu, \sigma^2)}{U(\theta_i \mid -a, a)}\right) d\theta$$

Breaking the integral into two parts:

$$D_{\mathrm{KL}}(x(\theta) \parallel y(\theta)) = \int_\theta \mathcal{N}(\theta \mid \mu, \sigma^2) \log\left(\mathcal{N}(\theta \mid \mu, \sigma^2)\right) d\theta - \int_{-a}^{a} \mathcal{N}(\theta \mid \mu, \sigma^2) \log\left(\frac{1}{2a}\right) d\theta$$

Evaluating each term separately:

$$\begin{aligned}\int_\theta \mathcal{N}(\theta \mid \mu, \sigma^2) \log\left(\mathcal{N}(\theta \mid \mu, \sigma^2)\right) d\theta &= \mathbb{E}\left[\log\left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2\right)\right)\right] \\ &= \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + \mathbb{E}\left[-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2\right] \\ &= \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\end{aligned}$$

And:

$$\begin{aligned}\int_{-a}^{a} \mathcal{N}(\theta \mid \mu, \sigma^2) \log\left(\frac{1}{2a}\right) d\theta &= \log\left(\frac{1}{2a}\right) \int_{-a}^{a} \mathcal{N}(\theta \mid \mu, \sigma^2) d\theta \\ &= \log\left(\frac{1}{2a}\right) \left(\int_{-\infty}^{a} \mathcal{N}(\theta \mid \mu, \sigma^2) d\theta - \int_{-\infty}^{-a} \mathcal{N}(\theta \mid \mu, \sigma^2) d\theta\right) \\ &= \log\left(\frac{1}{2a}\right) \left(\Phi\left(\frac{a-\mu}{\sigma}\right) - \Phi\left(\frac{-a-\mu}{\sigma}\right)\right)\end{aligned}$$

Combining these results:

$$\begin{aligned}D_{\mathrm{KL}}(x(\theta) \,\|\, y(\theta)) &= \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2} - \log\left(\frac{1}{2a}\right) \left(\Phi\left(\frac{a-\mu}{\sigma}\right) - \Phi\left(\frac{-a-\mu}{\sigma}\right)\right) \\ &= \log\left(\frac{1}{\sigma}\right) - \log\left(\sqrt{2\pi}\right) - \frac{1}{2} + \log(2a) \left(\Phi\left(\frac{a-\mu}{\sigma}\right) - \Phi\left(\frac{-a-\mu}{\sigma}\right)\right) \\ &= \log\left(\frac{1}{\sigma}\right) + \log(2a) \left(\Phi\left(\frac{a-\mu}{\sigma}\right) - \Phi\left(\frac{-a-\mu}{\sigma}\right)\right) - \left(\frac{1}{2} + \log\left(\sqrt{2\pi}\right)\right)\end{aligned}$$

Hence, we have proven the expression for $D_{\mathrm{KL}}(x(\theta) \,\|\, y(\theta))$ as desired.

□

Theorem 5 (KLD between marginal true prior and marginal assumed prior)**.** *The KLD $D_{KL}(p(\theta_i) \,\|\, q(\theta_i))$ between the marginal true prior $p(\theta_i)$ and the marginal assumed prior $q(\theta_i)$ is given by:*

$$D_{KL}(p(\theta_i) \,\|\, q(\theta_i)) = \log\left(\frac{1}{\sigma_p}\right) + \log(6) \left(\Phi\left(\frac{3-\mu_p}{\sigma_p}\right) - \Phi\left(\frac{-3-\mu_p}{\sigma_p}\right)\right) - \left(\frac{1}{2} + \log\left(\sqrt{2\pi}\right)\right)$$

where Φ denotes the cumulative distribution function (CDF) of the standard normal distribution.

Proof. Using the result from Lemma 1, we know the KLD between a normal distribution and a uniform distribution. Here, the marginal true prior $p(\theta_i)$ is a normal distribution $\mathcal{N}(\theta_i \mid \mu_{\mathrm{p}}, \sigma_{\mathrm{p}}^2)$ and the marginal assumed prior $q(\theta_i)$ is a uniform distribution over $[-3, 3]$.

Substituting $\mu = \mu_{\mathrm{p}}$, $\sigma = \sigma_{\mathrm{p}}$, and $a = 3$ into the lemma's result, we get:

$$D_{\mathrm{KL}}(p(\theta_i) \parallel q(\theta_i)) = \log\left(\frac{1}{\sigma_{\mathrm{p}}}\right) + \log(6)\left(\Phi\left(\frac{3-\mu_p}{\sigma_p}\right) - \Phi\left(\frac{-3-\mu_p}{\sigma_p}\right)\right) - \left(\frac{1}{2} + \log\left(\sqrt{2\pi}\right)\right)$$

Thus, we have proven the expression for $D_{\mathrm{KL}}(p(\theta_i) \parallel q(\theta_i))$ as desired.

□

Lemma 2 (The KLD of one normal distribution from another normal distribution)**.** *The Kullback–Leibler divergence of one normal distribution* $x(\theta) = \mathcal{N}(\theta \mid \mu_x, {\sigma_x}^2)$ *from another* $y(\theta) = \mathcal{N}(\theta \mid \mu_y, {\sigma_y}^2)$ *is given by*

$$D_{KL}(x(\theta) \parallel y(\theta)) = \log\left(\frac{\sigma_y}{\sigma_x}\right) + \frac{{\sigma_x}^2 + (\mu_x - \mu_y)^2}{2{\sigma_y}^2} - \frac{1}{2}$$

Proof. Starting with the definition of KLD, we have:

$$\begin{aligned} D_{\mathrm{KL}}(x(\theta) \parallel y(\theta)) &= \int_\theta x(\theta) \log\left(\frac{x(\theta)}{q(\theta)}\right) d\theta \\ &= \int_\theta x(\theta) \log(x(\theta))\, d\theta - \int_\theta x(\theta) \log(y(\theta))\, d\theta \\ &= \mathbb{E}\left[\log(x(\theta))\right]_{x(\theta)} - \mathbb{E}\left[\log(y(\theta))\right]_{x(\theta)} \end{aligned}$$

Substituting the given expressions for $x(\theta)$ and $y(\theta)$, we get:

$$
\begin{aligned}
D_{\mathrm{KL}}(x(\theta) \,\|\, y(\theta)) &= \mathbb{E}\left[\log\left(\mathcal{N}(\theta, \mu_x, \sigma_x)\right)\right]_{x(\theta)} - \mathbb{E}\left[\log\left(\mathcal{N}(\theta, \mu_y, \sigma_y)\right)\right]_{x(\theta)} \\
&= \mathbb{E}\left[\log\left(\frac{1}{\sigma_x\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\theta-\mu_x}{\sigma_x}\right)^2\right)\right)\right]_{x(\theta)} - \mathbb{E}\left[\log\left(\frac{1}{\sigma_y\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\theta-\mu_y}{\sigma_y}\right)^2\right)\right)\right]_{x(\theta)} \\
&= \mathbb{E}\left[\log\left(\frac{1}{\sigma_x\sqrt{2\pi}}\right)\right]_{x(\theta)} + \mathbb{E}\left[\log\left(\exp\left(-\frac{1}{2}\left(\frac{\theta-\mu_x}{\sigma_x}\right)^2\right)\right)\right]_{x(\theta)} \\
&\quad - \mathbb{E}\left[\log\left(\frac{1}{\sigma_y\sqrt{2\pi}}\right)\right]_{x(\theta)} - \mathbb{E}\left[\log\left(\exp\left(-\frac{1}{2}\left(\frac{\theta-\mu_y}{\sigma_y}\right)^2\right)\right)\right]_{x(\theta)} \\
&= \log\left(\frac{1}{\sigma_x\sqrt{2\pi}}\right) - \log\left(\frac{1}{\sigma_y\sqrt{2\pi}}\right) + \mathbb{E}\left[\left(-\frac{1}{2}\left(\frac{\theta-\mu_x}{\sigma_x}\right)^2\right)\right]_{x(\theta)} - \mathbb{E}\left[\left(-\frac{1}{2}\left(\frac{\theta-\mu_y}{\sigma_y}\right)^2\right)\right]_{x(\theta)} \\
&= \log\left(\frac{\sigma_y}{\sigma_x}\right) - \frac{1}{2{\sigma_x}^2}\mathbb{E}\left[(\theta-\mu_x)^2\right]_{x(\theta)} + \frac{1}{2{\sigma_y}^2}\mathbb{E}\left[(\theta-\mu_y)^2\right]_{x(\theta)} \\
&= \log\left(\frac{\sigma_y}{\sigma_x}\right) - \frac{{\sigma_x}^2}{2{\sigma_x}^2} + \frac{\mathbb{E}\left[\theta^2 - 2\theta\mu_y + {\mu_y}^2\right]_{x(\theta)}}{2{\sigma_y}^2} \\
&= \log\left(\frac{\sigma_y}{\sigma_x}\right) - \frac{1}{2} + \frac{\mathbb{E}\left[\theta^2\right]_{x(\theta} - 2\mathbb{E}\left[\theta\right]_{x(\theta)}\mu_y + {\mu_y}^2}{2{\sigma_y}^2} \\
&= \log\left(\frac{\sigma_y}{\sigma_x}\right) + \frac{\mathbb{E}\left[\theta^2\right]_{x(\theta} - 2\mu_x\mu_y + {\mu_y}^2}{2{\sigma_y}^2} - \frac{1}{2} \\
&= \log\left(\frac{\sigma_y}{\sigma_x}\right) + \frac{\left(\mathbb{E}\left[\theta^2\right]_{x(\theta} - \mu_x^2\right) + \left(\mu_x^2 - 2\mu_x\mu_y + {\mu_y}^2\right)}{2{\sigma_y}^2} - \frac{1}{2} \\
&= \log\left(\frac{\sigma_y}{\sigma_x}\right) + \frac{\sigma_x^2 + (\mu_x - \mu_y)^2}{2{\sigma_y}^2} - \frac{1}{2}
\end{aligned}
$$

Hence, we have proven the expression for $D_{\mathrm{KL}}(x(\theta) \,\|\, y(\theta))$ as desired. □

Theorem 6 (KLD between approximated marginal true posterior and marginal estimated posterior)**.** *The KLD $D_{KL}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c))$ between the approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ and the marginal estimated posterior $q(\theta_i \mid c)$ is given by:*

$$
D_{KL}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) = \log\left(\frac{\sigma_\ell}{\sigma_\phi}\right) + \frac{{\sigma_\phi}^2 + (\mu_\phi - \mu_\ell)^2}{2{\sigma_\ell}^2} - \frac{1}{2}
$$

Proof. To prove Theorem 6, we'll use Lemma 2 to find the KLD between the approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ and the marginal estimated posterior $q(\theta_i \mid c)$, both of which are normal distributions.

From Lemma 2, we have:

$$
D_{\mathrm{KL}}(x(\theta) \,\|\, y(\theta)) = \log\left(\frac{\sigma_y}{\sigma_x}\right) + \frac{{\sigma_x}^2 + (\mu_x - \mu_y)^2}{2{\sigma_y}^2} - \frac{1}{2}
$$

Let $x(\theta) = \hat{p}(\theta_i \mid c)$ and $y(\theta) = q(\theta_i \mid c)$.

According to Theorem 3, the marginal estimated posterior $q(\theta_i \mid c)$ has mean μ_ℓ and variance σ_ℓ^2:

$$q(\theta_i \mid c) = \mathcal{N}(\theta_i \mid \mu_\ell, \sigma_\ell^2)$$

Similarly, according to Theorem 4, the approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ has mean μ_ϕ and variance σ_ϕ^2:

$$\hat{p}(\theta_i \mid c) = \mathcal{N}(\theta_i \mid \mu_\phi, \sigma_\phi^2)$$

Substituting these expressions into the formula for the KLD, we get:

$$D_{\mathrm{KL}}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) = \log\left(\frac{\sigma_\ell}{\sigma_\phi}\right) + \frac{{\sigma_\phi}^2 + (\mu_\phi - \mu_\ell)^2}{2{\sigma_\ell}^2} - \frac{1}{2}$$

Therefore, the KLD between the approximated marginal true posterior and the marginal estimated posterior is given by the expression provided in Theorem 6, as desired.

□

Lemma 3 (Derivative of the standard normal CDF with respect to standard deviation)**.** *Let Φ denote the cumulative distribution function (CDF) of the standard normal distribution, and let ϕ denote the probability density function (PDF) of the standard normal distribution. For a normal distribution with mean μ and standard deviation σ, the derivative of the CDF Φ with respect to the standard deviation σ is given by:*

$$\frac{\partial}{\partial\sigma}\Phi\left(\frac{\theta-\mu}{\sigma}\right) = -\left(\frac{\theta-\mu}{\sigma^2}\right)\phi\left(\frac{\theta-\mu}{\sigma}\right)$$

Proof. To prove this lemma, we start by noting the definitions of Φ and ϕ:

$$\Phi(z) = \int_{-\infty}^{z} \phi(t)\,dt, \quad \text{where} \quad \phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Let $z = \frac{\theta-\mu}{\sigma}$. Then we have:

$$\Phi\left(\frac{\theta-\mu}{\sigma}\right) = \Phi(z)$$

Taking the partial derivative of $\Phi(z)$ with respect to σ:

$$\frac{\partial}{\partial\sigma}\Phi\left(\frac{\theta-\mu}{\sigma}\right) = \frac{\partial\Phi}{\partial z}\cdot\frac{\partial z}{\partial\sigma}$$

Since $\frac{\partial\Phi}{\partial z} = \phi(z)$, we have:

$$\frac{\partial\Phi}{\partial z} = \phi\left(\frac{\theta-\mu}{\sigma}\right)$$

Next, we calculate $\frac{\partial z}{\partial \sigma}$:

$$z = \frac{\theta - \mu}{\sigma} \implies \frac{\partial z}{\partial \sigma} = \frac{\partial}{\partial \sigma}\left(\frac{\theta - \mu}{\sigma}\right) = -\frac{\theta - \mu}{\sigma^2}$$

Combining these results, we obtain:

$$\frac{\partial}{\partial \sigma}\Phi\left(\frac{\theta - \mu}{\sigma}\right) = \phi\left(\frac{\theta - \mu}{\sigma}\right) \cdot \left(-\frac{\theta - \mu}{\sigma^2}\right)$$

Simplifying, we get the desired result:

$$\frac{\partial}{\partial \sigma}\Phi\left(\frac{\theta - \mu}{\sigma}\right) = -\left(\frac{\theta - \mu}{\sigma^2}\right)\phi\left(\frac{\theta - \mu}{\sigma}\right)$$

Thus, the lemma is proven.

□

Theorem 7 (Derivative of KLD between marginal true prior and marginal assumed prior with respect to standard deviation)**.** *The partial derivative of the KLD $D_{KL}(p(\theta_i) \,\|\, q(\theta_i))$ between the marginal true prior $p(\theta_i)$ and the marginal assumed prior $q(\theta_i)$ with respect to the standard deviation σ_p of the marginal true prior is:*

$$\frac{\partial}{\partial \sigma_p} D_{KL}(p(\theta_i) \,\|\, q(\theta_i)) = -\frac{1}{\sigma_p} - \log(6)\left(\left(\frac{3 - \mu_p}{\sigma_p^2}\right)\phi\left(\frac{3 - \mu_p}{\sigma_p}\right) + \left(\frac{3 + \mu_p}{\sigma_p^2}\right)\phi\left(\frac{-3 - \mu_p}{\sigma_p}\right)\right)$$

where ϕ denotes the probability density function (PDF) of the standard normal distribution.

Proof. To prove Theorem 7, we will start by differentiating the expression for the KLD given in Theorem 5 with respect to the standard deviation σ_p.

Given:

$$D_\mathrm{KL}(p(\theta_i) \,\|\, q(\theta_i)) = \log\left(\frac{1}{\sigma_\mathrm{p}}\right) + \log(6)\left(\Phi\left(\frac{3 - \mu_p}{\sigma_p}\right) - \Phi\left(\frac{-3 - \mu_p}{\sigma_p}\right)\right) - \left(\frac{1}{2} + \log\left(\sqrt{2\pi}\right)\right)$$

We differentiate each term separately with respect to σ_p.

1. The derivative of the first term:
$$\frac{\partial}{\partial \sigma_\mathrm{p}} \log\left(\frac{1}{\sigma_\mathrm{p}}\right) = -\frac{1}{\sigma_\mathrm{p}}$$

2. The derivative of the second term:

$$
\begin{aligned}
&\frac{\partial}{\partial \sigma_{\mathrm{p}}}\left(\log(6)\left(\Phi\left(\frac{3-\mu_p}{\sigma_p}\right)-\Phi\left(\frac{-3-\mu_p}{\sigma_p}\right)\right)\right)\\
&=\log(6)\left(\frac{\partial}{\partial \sigma_{\mathrm{p}}}\Phi\left(\frac{3-\mu_p}{\sigma_p}\right)-\frac{\partial}{\partial \sigma_{\mathrm{p}}}\Phi\left(\frac{-3-\mu_p}{\sigma_p}\right)\right)\\
&=\log(6)\left(-\left(\frac{3-\mu_p}{\sigma_p^2}\right)\phi\left(\frac{3-\mu_p}{\sigma_p}\right)-\left(\frac{3+\mu_p}{\sigma_p^2}\right)\phi\left(\frac{-3-\mu_p}{\sigma_p}\right)\right) \qquad \text{(using Lemma 3)}\\
&=-\log(6)\left(\left(\frac{3-\mu_p}{\sigma_p^2}\right)\phi\left(\frac{3-\mu_p}{\sigma_p}\right)+\left(\frac{3+\mu_p}{\sigma_p^2}\right)\phi\left(\frac{-3-\mu_p}{\sigma_p}\right)\right)
\end{aligned}
$$

3. The derivative of the third term:

$$
\frac{\partial}{\partial \sigma_{\mathrm{p}}}\left(-\left(\frac{1}{2}+\log\left(\sqrt{2\pi}\right)\right)\right)=0
$$

Combining these results, we obtain:

$$
\frac{\partial}{\partial \sigma_{\mathrm{p}}} D_{\mathrm{KL}}(p(\theta_i)\,\|\,q(\theta_i))=-\frac{1}{\sigma_{\mathrm{p}}}-\log(6)\left(\left(\frac{3-\mu_p}{\sigma_p^2}\right)\phi\left(\frac{3-\mu_p}{\sigma_p}\right)+\left(\frac{3+\mu_p}{\sigma_p^2}\right)\phi\left(\frac{-3-\mu_p}{\sigma_p}\right)\right)
$$

Thus, the derivative of the KLD between the marginal true prior $p(\theta_i)$ and the marginal assumed prior $q(\theta_i)$ with respect to the standard deviation σ_{p} is derived as desired.

□

Theorem 8 (Derivative of KLD between approximated marginal true posterior and marginal estimated posterior)**.** *The derivative of the KLD between the approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ and the marginal estimated posterior $q(\theta_i \mid c)$ with respect to the standard deviation σ_p of the marginal true prior is:*

$$
\frac{\partial}{\partial \sigma_p} D_{KL}(\hat{p}(\theta_i \mid c)\,\|\,q(\theta_i \mid c))=-\frac{{\sigma_\ell}^2}{{\sigma_\ell}^2\sigma_p+{\sigma_p}^3}-\frac{{\sigma_\ell}^2\sigma_p\left(2(\mu_p-\mu_\ell)^2-\left({\sigma_\ell}^2+{\sigma_p}^2\right)\right)}{({\sigma_\ell}^2+{\sigma_p}^2)^3}
$$

Proof. To prove Theorem 8, we will first rewrite the Kullback-Leibler divergence (KLD) $D_{\mathrm{KL}}(\hat{p}(\theta_i \mid c)\,\|\,q(\theta_i \mid c))$ in terms of σ_{p}, and then differentiate it with respect to σ_{p}.

Given Theorem 6, we have:

$$
D_{\mathrm{KL}}(\hat{p}(\theta_i \mid c)\,\|\,q(\theta_i \mid c))=\log\left(\frac{\sigma_\ell}{\sigma_\phi}\right)+\frac{{\sigma_\phi}^2+(\mu_\phi-\mu_\ell)^2}{2{\sigma_\ell}^2}-\frac{1}{2}
$$

Now, expressing σ_ϕ in terms of σ_{p} using the formulas given in Theorem 4:

$$
\mu_\phi=\frac{\mu_{\mathrm{p}}\sigma_\ell^2+\mu_\ell\sigma_{\mathrm{p}}^2}{\sigma_\ell^2+\sigma_{\mathrm{p}}^2}
$$

$$
\sigma_\phi^2=\frac{\sigma_{\mathrm{p}}^2\sigma_\ell^2}{\sigma_{\mathrm{p}}^2+\sigma_\ell^2}
$$

We substitute μ_ϕ and σ_ϕ into the expression for D_{KL}:

$$D_{\mathrm{KL}}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) = \log\left(\frac{\sigma_\ell}{\sqrt{\frac{\sigma_{\mathrm{p}}^2\sigma_\ell^2}{\sigma_{\mathrm{p}}^2+\sigma_\ell^2}}}\right) + \frac{\frac{\sigma_{\mathrm{p}}^2\sigma_\ell^2}{\sigma_{\mathrm{p}}^2+\sigma_\ell^2} + (\frac{\mu_{\mathrm{p}}\sigma_\ell^2+\mu_\ell\sigma_{\mathrm{p}}^2}{\sigma_\ell^2+\sigma_{\mathrm{p}}^2} - \mu_\ell)^2}{2{\sigma_\ell}^2} - \frac{1}{2}$$

Simplify the expression:

$$D_{\mathrm{KL}}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) = \log\left(\frac{\sigma_\ell\sqrt{\sigma_{\mathrm{p}}^2+\sigma_\ell^2}}{\sigma_{\mathrm{p}}}\right) + \frac{\sigma_{\mathrm{p}}^2 + \frac{\sigma_\ell^2(\mu_{\mathrm{p}}-\mu_\ell)^2}{\sigma_{\mathrm{p}}^2+\sigma_\ell^2}}{2(\sigma_{\mathrm{p}}^2+\sigma_\ell^2)} - \frac{1}{2}$$

Now, we differentiate D_{KL} with respect to σ_{p}. The derivative of the first term involves the chain rule, and the derivative of the second term can be obtained directly:

$$\frac{\partial}{\partial\sigma_{\mathrm{p}}} D_{\mathrm{KL}}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) = \frac{\partial}{\partial\sigma_{\mathrm{p}}}\left(\log\left(\frac{\sigma_\ell\sqrt{\sigma_{\mathrm{p}}^2+\sigma_\ell^2}}{\sigma_{\mathrm{p}}}\right)\right) + \frac{\partial}{\partial\sigma_{\mathrm{p}}}\left(\frac{\sigma_{\mathrm{p}}^2 + \frac{\sigma_\ell^2(\mu_{\mathrm{p}}-\mu_\ell)^2}{\sigma_{\mathrm{p}}^2+\sigma_\ell^2}}{2(\sigma_{\mathrm{p}}^2+\sigma_\ell^2)}\right)$$

After computing the derivatives, we arrive at the expression:

$$\frac{\partial}{\partial\sigma_{\mathrm{p}}} D_{\mathrm{KL}}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) = -\frac{{\sigma_\ell}^2}{{\sigma_\ell}^2\sigma_{\mathrm{p}} + {\sigma_{\mathrm{p}}}^3} - \frac{{\sigma_\ell}^2\sigma_{\mathrm{p}}\left(2(\mu_{\mathrm{p}}-\mu_\ell)^2 - \left({\sigma_\ell}^2+{\sigma_{\mathrm{p}}}^2\right)\right)}{({\sigma_\ell}^2+{\sigma_{\mathrm{p}}}^2)^3}$$

This completes the proof of Theorem 8. □

Lemma 4 (0.5 Points of a Likelihood)**.** *Consider a likelihood $P(c=1 \mid \theta_i)$ that follows a univariate Gaussian distribution:*

$$P(c=1 \mid \theta_i) = \mathcal{N}(\theta_i \mid \mu, \sigma^2)$$

The 0.5 points are the values of θ_i for which the likelihood function equals 0.5. The left and right 0.5 points of this distribution are given by:

$$\theta_{left} = \mu - \sigma\sqrt{2\log\left(\frac{2}{\sigma\sqrt{2\pi}}\right)}$$

$$\theta_{right} = \mu + \sigma\sqrt{2\log\left(\frac{2}{\sigma\sqrt{2\pi}}\right)}$$

Proof. For a likelihood that follows a Gaussian distribution $\mathcal{N}(\theta_i \mid \mu, \sigma^2)$, the function is given by:

$$f(\theta_i) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\theta_i-\mu)^2}{2\sigma^2}\right)$$

The 0.5 points occur where the likelihood function equals 0.5:

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\theta_{\text{left}}-\mu)^2}{2\sigma^2}\right) = \frac{1}{2}$$

Taking the natural logarithm of both sides:

$$\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(\theta_{\text{left}} - \mu)^2}{2\sigma^2} = \log\left(\frac{1}{2}\right)$$

Simplifying the right-hand side:

$$-\frac{(\theta_{\text{left}} - \mu)^2}{2\sigma^2} = \log\left(\frac{1}{2}\right) - \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)$$

$$-\frac{(\theta_{\text{left}} - \mu)^2}{2\sigma^2} = -\log 2 - \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)$$

$$\frac{(\theta_{\text{left}} - \mu)^2}{2\sigma^2} = \log\left(\frac{2}{\sigma\sqrt{2\pi}}\right)$$

Solving for θ_{left}:

$$(\theta_{\text{left}} - \mu)^2 = 2\sigma^2 \log\left(\frac{2}{\sigma\sqrt{2\pi}}\right)$$

$$\theta_{\text{left}} - \mu = \pm\sigma\sqrt{2\log\left(\frac{2}{\sigma\sqrt{2\pi}}\right)}$$

Since θ_{left} is to the left of μ:

$$\theta_{\text{left}} = \mu - \sigma\sqrt{2\log\left(\frac{2}{\sigma\sqrt{2\pi}}\right)}$$

Similarly, for the right 0.5 point:

$$\theta_{\text{right}} = \mu + \sigma\sqrt{2\log\left(\frac{2}{\sigma\sqrt{2\pi}}\right)}$$

Thus, we have derived the desired formulas for the left and right 0.5 points of the likelihood function. □

Theorem 9 (Marginal estimated posterior via inaccurate likelihood)**.** *The marginal estimated posterior $q(\theta_i \mid c)$ for the parameter θ_i via inaccurate likelihood $q(\theta_i \mid c) = \mathcal{N}(\theta_i \mid \mu_q, \sigma_q^2)$ is a univariate Gaussian distribution with mean μ_q and covariance σ_ℓ:*

$$q(\theta_i \mid c) = \mathcal{N}(\theta \mid \mu_q, {\sigma_\ell}^2) \text{ for } -3 \leq \theta_i \leq 3$$

Proof. From Definition 12, we have that the marginal inaccurate likelihood $q(c \mid \theta_i)$ follows a Gaussian distribution with mean μ_q and variance σ_ℓ^2, i.e.,

$$q(c \mid \theta_i) = \mathcal{N}(\theta \mid \mu_q, {\sigma_\ell}^2)$$

From Theorem 1, we know that the marginal posterior distribution $q(\theta_i \mid c)$ given class c and a uniform prior $q(\theta_i)$ is proportional to the product of the marginal likelihood $q(c \mid \theta_i)$ and the marginal prior $q(\theta_i)$, i.e.,

$$q(\theta_i \mid c) \propto q(c \mid \theta_i) q(\theta_i) \text{ for } -3 \leq \theta_i \leq 3$$

Substituting the expression for $p(c \mid \theta_i)$ into the proportionality, we get:

$$q(\theta_i \mid c) \propto \mathcal{N}(\theta \mid \mu_q, {\sigma_\ell}^2) \cdot q(\theta_i) \text{ for } -3 \leq \theta_i \leq 3$$

Since $q(\theta_i)$ is a uniform distribution, it is constant, and hence, we can drop the proportionality sign. Therefore,

$$q(\theta_i \mid c) = \mathcal{N}(\theta \mid \mu_q, {\sigma_\ell}^2) \text{ for } -3 \leq \theta_i \leq 3$$

Thus, the theorem is proved.

□

Theorem 10 (Error generated by marginal inaccurate likelihood)**.** *The error generated by this model, represented as the cumulative probability of false positives (FP) and false negatives (FN), arises due to the discrepancy $\beta = |\mu_\ell - \mu_q|$ between μ_q and μ_ℓ.*

Case 1: *$\mu_q = \mu_\ell - \beta$ (This case is depicted in Figure B.1 (a))*

$$FP_1 = \int_{\theta_{left}}^{\theta_{left}+\beta} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i$$

$$FN_1 = \int_{\theta_{right}}^{\theta_{right}+\beta} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i$$

Case 2: *$\mu_q = \mu_\ell + \beta$ (This case is depicted in Figure B.1 (b))*

$$FN_2 = \int_{\theta'_{left}}^{\theta'_{left}+\beta} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i$$

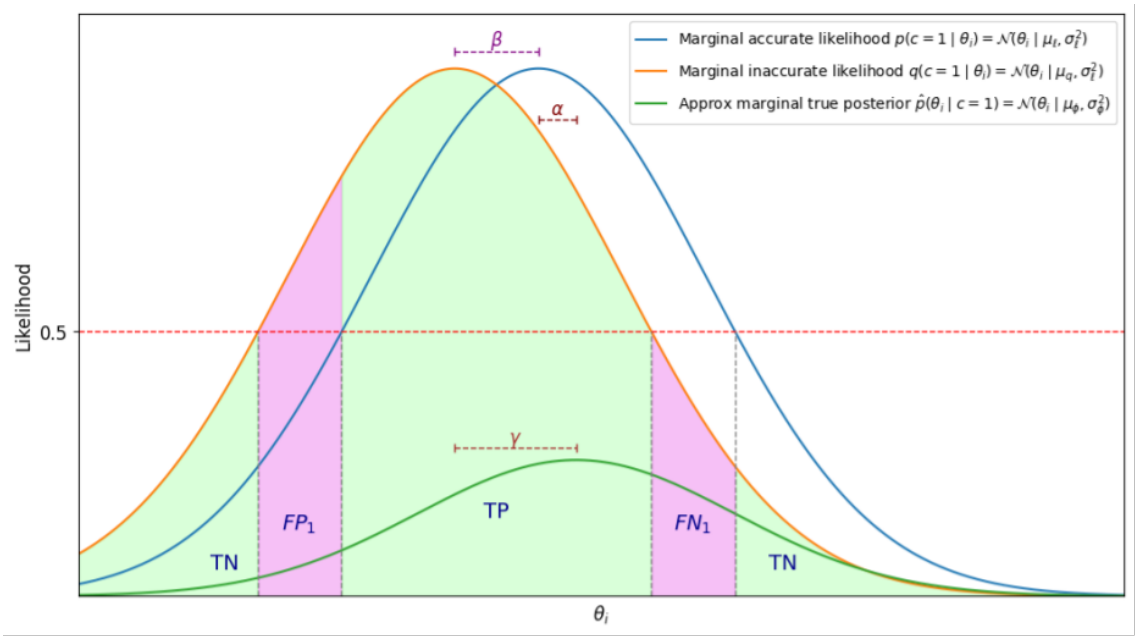
$$FP_2 = \int_{\theta'_{right}}^{\theta'_{right}+\beta} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i$$

where θ_{left} and θ_{right} are the left and right 0.5 points of the marginal inaccurate likelihood, respectively; while θ'_{left} and θ'_{right} are the left and right 0.5 points of the marginal accurate likelihood, respectively.

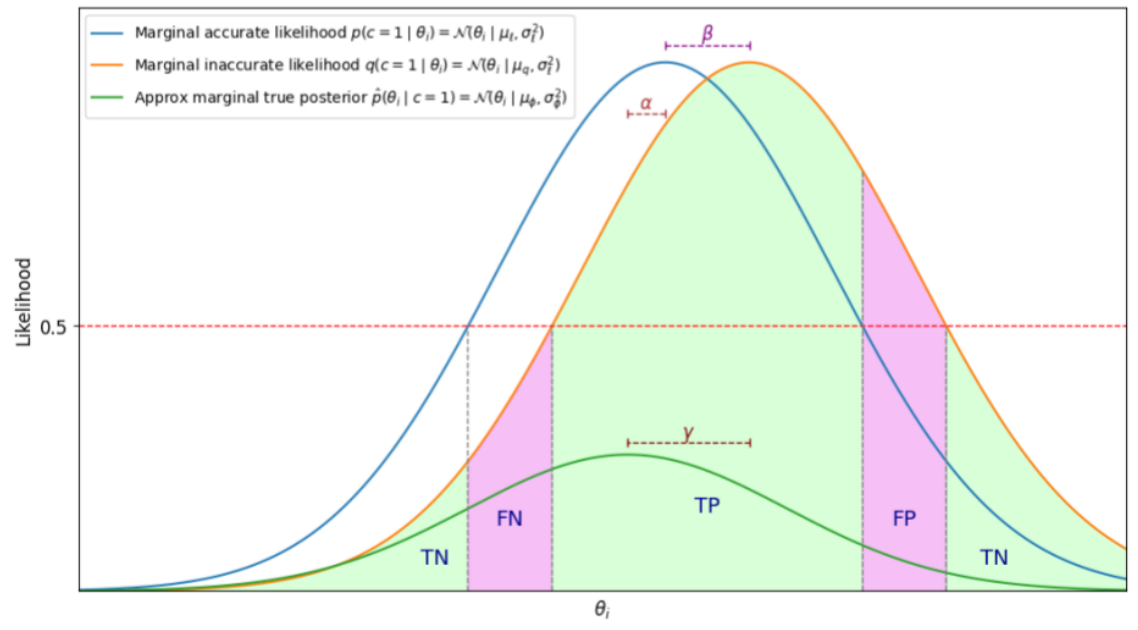
Proof. For the marginal inaccurate likelihood that follows a Gaussian function $\mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)$, the left and right 0.5 points are symmetric around μ_q. According to Lemma 4, these points are located at:

$$\theta_{\text{left}} = \mu_q - \sigma_\ell \sqrt{2 \log\left(\frac{2}{\sigma_\ell \sqrt{2\pi}}\right)}$$

$$\theta_{\text{right}} = \mu_q + \sigma_\ell \sqrt{2 \log\left(\frac{2}{\sigma_\ell \sqrt{2\pi}}\right)}$$



(a)



(b)

Figure B.1: **Visual comparison of marginal accurate and inaccurate likelihoods with classification outcomes.** This figure illustrates the marginal accurate likelihood $p(c \mid \theta_i)$ and the marginal inaccurate likelihood $q(c \mid \theta_i)$ for the parameter θ_i. In scenario (a), the marginal inaccurate likelihood $q(c \mid \theta_i)$ is shifted by $-\beta$ from the accurate likelihood $p(c \mid \theta_i)$, and in scenario (b), it is shifted by $+\beta$. The absolute mean differences α, β, and γ between these distributions are indicated, reflecting their respective relationships. The shaded areas under the curves represent different classification outcomes (True Positive, False Positive, True Negative, and False Negative) relative to the decision threshold of 0.5.

For the marginal accurate likelihood that follows a Gaussian function $\mathcal{N}(\theta_i \mid \mu_\ell, \sigma_\ell^2)$, the left and right 0.5 points are according to Lemma 4 located at:

$$\theta'_{\text{left}} = \mu_\ell - \sigma_\ell \sqrt{2 \log\left(\frac{2}{\sigma_\ell \sqrt{2\pi}}\right)}$$

$$\theta'_{\text{right}} = \mu_\ell + \sigma_\ell \sqrt{2 \log\left(\frac{2}{\sigma_\ell \sqrt{2\pi}}\right)}$$

To prove this theorem, we consider the two cases separately:

- Case 1: $\mu_q = \mu_\ell - \beta$

  In this case,

$$\begin{aligned}
\theta'_{\text{left}} &= \mu_\ell - \sigma_\ell \sqrt{2 \log\left(\frac{2}{\sigma_\ell \sqrt{2\pi}}\right)} \\
&= \mu_q + \beta - \sigma_\ell \sqrt{2 \log\left(\frac{2}{\sigma_\ell \sqrt{2\pi}}\right)} \\
&= \left(\mu_q - \sigma_\ell \sqrt{2 \log\left(\frac{2}{\sigma_\ell \sqrt{2\pi}}\right)}\right) + \beta \\
&= \theta_{\text{left}} + \beta
\end{aligned}$$

  The same way:

$$\theta'_{\text{right}} = \theta_{\text{right}} + \beta$$

  False positives (FP):

  FP occurs between the left 0.5 points of the two likelihood functions. In this area, the marginal inaccurate likelihood is already > 0.5 and thus the model predicts positive, but the marginal accurate likelihood is still < 0.5 suggesting the actual class is negative; hence FP. The area of FP is:

$$\text{FP}_1 = \int_{\theta_{\text{left}}}^{\theta'_{\text{left}}} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i = \int_{\theta_{\text{left}}}^{\theta_{\text{left}}+\beta} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i$$

  False negatives (FN):

  FN occurs between the right 0.5 points of the two likelihood functions. In this area, the marginal inaccurate likelihood is already < 0.5 and thus the model predicts negative, but the marginal accurate likelihood is still > 0.5 suggesting the actual class is positive; hence FN. The area of FN is:

$$\text{FN}_1 = \int_{\theta_{\text{right}}}^{\theta'_{\text{right}}} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i = \int_{\theta_{\text{right}}}^{\theta_{\text{right}}+\beta} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i$$

- Case 2: $\mu_q = \mu_\ell + \beta$

  In this case,

$$\begin{aligned}\theta_{\text{left}} &= \mu_q - \sigma_\ell \sqrt{2 \log\left(\frac{2}{\sigma_\ell \sqrt{2\pi}}\right)} \\ &= \mu_\ell + \beta - \sigma_\ell \sqrt{2 \log\left(\frac{2}{\sigma_\ell \sqrt{2\pi}}\right)} \\ &= \left(\mu_\ell - \sigma_\ell \sqrt{2 \log\left(\frac{2}{\sigma_\ell \sqrt{2\pi}}\right)}\right) + \beta \\ &= \theta'_{\text{left}} + \beta\end{aligned}$$

The same way:

$$\theta_{\text{right}} = \theta'_{\text{right}} + \beta$$

False negatives (FN):

FN occurs between the left 0.5 points of the two likelihood functions. In this area, the marginal inaccurate likelihood is still < 0.5 and thus the model predicts negative, but the marginal accurate likelihood is already > 0.5 suggesting the actual class is positive; hence FN. The area of FN is:

$$\text{FN}_2 = \int_{\theta'_{\text{left}}}^{\theta_{\text{left}}} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i = \int_{\theta'_{\text{left}}}^{\theta'_{\text{left}}+\beta} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i$$

False positives (FP):

FP occurs between the right 0.5 points of the two likelihood functions. In this area, the marginal inaccurate likelihood is still > 0.5 and thus the model predicts positive, but the marginal accurate likelihood is already < 0.5 suggesting the actual class is negative; hence FP. The area of FP is:

$$\text{FP}_2 = \int_{\theta'_{\text{right}}}^{\theta_{\text{right}}} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i = \int_{\theta'_{\text{right}}}^{\theta'_{\text{right}}+\beta} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i$$

Thus, we have derived the desired formulas for the false positive and false negative areas,

□

Theorem 11 (KLD between approximated marginal true posterior and marginal estimated posterior via inaccurate likelihood)**.** *The KLD $D_{KL}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c))$ between the approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ and the marginal estimated posterior $q(\theta_i \mid c)$ via inaccurate likelihood $q(\theta_i \mid c) = \mathcal{N}(\theta_i \mid \mu_q, \sigma_q^2)$ is given by:*

$$D_{KL}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) = \log\left(\frac{\sigma_q}{\sigma_\phi}\right) + \frac{{\sigma_\phi}^2 + (\alpha+\beta)^2}{2\sigma_q^2} - \frac{1}{2}$$

Proof. To prove Theorem 11, we will use Lemma 2 to find the KLD between the approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ and the marginal estimated posterior $q(\theta_i \mid c)$ via inaccurate likelihood $q(\theta_i \mid c)$, both of which are normal distributions.

From Proposition 2, we have:

$$D_{\text{KL}}(x(\theta) \,\|\, y(\theta)) = \log\left(\frac{\sigma_y}{\sigma_x}\right) + \frac{{\sigma_x}^2 + (\mu_x - \mu_y)^2}{2{\sigma_y}^2} - \frac{1}{2}$$

Let $x(\theta) = \hat{p}(\theta_i \mid c)$ and $y(\theta) = q(\theta_i \mid c)$.

According to Theorem 4, the approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ has mean μ_ϕ and variance σ_ϕ^2:

$$\hat{p}(\theta_i \mid c) = \mathcal{N}(\theta_i \mid \mu_\phi, \sigma_\phi^2)$$

Similarly, according to Theorem 9, the marginal estimated posterior $q(\theta_i \mid c)$ via inaccurate likelihood has mean μ_q and variance σ_ℓ^2:

$$q(\theta_i \mid c) = \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)$$

Substituting these expressions into the formula for the KLD, we get:

$$\begin{aligned} D_{\mathrm{KL}}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) &= \log\left(\frac{\sigma_\ell}{\sigma_\phi}\right) + \frac{{\sigma_\phi}^2 + (\mu_\phi - \mu_q)^2}{2{\sigma_\ell}^2} - \frac{1}{2} \\ &= \log\left(\frac{\sigma_\ell}{\sigma_\phi}\right) + \frac{{\sigma_\phi}^2 + |\mu_\phi - \mu_q|^2}{2{\sigma_\ell}^2} - \frac{1}{2} \end{aligned}$$

As Definition 12 suggests $\gamma = \alpha + \beta = |\mu_\phi - \mu_q|$,

$$D_{\mathrm{KL}}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) = \log\left(\frac{\sigma_q}{\sigma_\phi}\right) + \frac{{\sigma_\phi}^2 + (\alpha + \beta)^2}{2\sigma_q^2} - \frac{1}{2}$$

Therefore, the KLD between the approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ and the marginal estimated posterior $q(\theta_i \mid c)$ via inaccurate likelihood $q(\theta_i \mid c)$ is given by the expression provided in Theorem 11, as desired.

□

Theorem 12 (Derivative of error with respect to β)**.** *The derivative of the error E (represented as false positives or false negatives given in Theorem 10) with respect to the absolute mean difference β between the marginal accurate likelihood and the marginal inaccurate likelihood is given by:*

$$\frac{\partial}{\partial \beta} E = \mathcal{N}(\theta_i = u \mid \mu_q, \sigma_\ell^2)$$

where u denotes the upper bound of the integral of the respective error area given in Theorem 10.

Proof. According to Theorem 10, the error area E is represented as either FP_1, FN_1, FN_2, or FP_2.

When E denotes FP_1, the derivative of the false positive area with respect to β is given by:

$$\frac{\partial}{\partial \beta} \mathrm{FP}_1 = \frac{\partial}{\partial \beta} \int_{\theta_{\mathrm{left}}}^{\theta_{\mathrm{left}}+\beta} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i$$

Substitute u for the upper bound of the integral $(\theta_{\mathrm{left}} + \beta)$:

$$\frac{\partial}{\partial\beta}\mathrm{FP}_1 = \frac{d}{du}\int_{\theta_{\text{left}}}^{u} \mathcal{N}(\theta_i \mid \mu_q, \sigma_\ell^2)\, d\theta_i \times \frac{du}{d\beta}$$

Since θ_{left} is a constant and $\frac{du}{d\beta} = 1$:

$$\frac{\partial}{\partial\beta}\mathrm{FP}_1 = \frac{\partial}{\partial\beta}\mathrm{E} = \mathcal{N}(\theta_i = u \mid \mu_q, \sigma_\ell^2)$$

Similarly, when E denotes either FN_1, FN_2, or FP_2, by substituting u for the upper bound of the respective integral:

$$\frac{\partial}{\partial\beta}\mathrm{E} = \mathcal{N}(\theta_i = u \mid \mu_q, \sigma_\ell^2)$$

Thus, we have derived the desired formula for the derivative of the error with respect to β.

□

Theorem 13 (Derivative of KLD between approximated marginal true posterior and marginal estimated posterior with respect to β)**.** *The derivative of the KLD $D_{KL}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c))$ between the approximated marginal true posterior $\hat{p}(\theta_i \mid c)$ and the marginal estimated posterior $q(\theta_i \mid c)$ via inaccurate likelihood $q(\theta_i \mid c) = \mathcal{N}(\theta_i \mid \mu_q, \sigma_q^2)$ with respect to the absolute mean difference between the accurate likelihood and the inaccurate likelihood β is given by:*

$$\frac{\partial}{\partial\beta} D_{KL}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) = \frac{\alpha + \beta}{\sigma_q^2}$$

Proof. From Theorem 11, we have the KLD between the approximated marginal true posterior and the marginal estimated posterior via inaccurate likelihood as:

$$D_{\mathrm{KL}}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) = \log\left(\frac{\sigma_q}{\sigma_\phi}\right) + \frac{{\sigma_\phi}^2 + (\alpha + \beta)^2}{2\sigma_q^2} - \frac{1}{2}$$

To find the derivative of this KLD with respect to β, we differentiate the right-hand side with respect to β:

$$\frac{\partial}{\partial\beta} D_{\mathrm{KL}}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) = \frac{\partial}{\partial\beta}\left(\log\left(\frac{\sigma_q}{\sigma_\phi}\right) + \frac{{\sigma_\phi}^2 + (\alpha + \beta)^2}{2\sigma_q^2} - \frac{1}{2}\right)$$

Since σ_q and σ_ϕ are constants with respect to β, the derivative of the logarithmic term is zero. Thus, we focus on the second term:

$$\frac{\partial}{\partial\beta}\left(\frac{{\sigma_\phi}^2 + (\alpha + \beta)^2}{2\sigma_q^2}\right)$$

Using the chain rule:

$$\frac{\partial}{\partial \beta}\left(\frac{{\sigma_\phi}^2 + (\alpha+\beta)^2}{2\sigma_q^2}\right) = \frac{1}{2\sigma_q^2} \cdot \frac{\partial}{\partial \beta}\left({\sigma_\phi}^2 + (\alpha+\beta)^2\right)$$

The term ${\sigma_\phi}^2$ is constant with respect to β, so its derivative is zero. Therefore:

$$\frac{\partial}{\partial \beta}\left({\sigma_\phi}^2 + (\alpha+\beta)^2\right) = \frac{\partial}{\partial \beta}\left((\alpha+\beta)^2\right) = 2(\alpha+\beta)$$

Substituting this back in:

$$\frac{\partial}{\partial \beta}\left(\frac{{\sigma_\phi}^2 + (\alpha+\beta)^2}{2\sigma_q^2}\right) = \frac{1}{2\sigma_q^2} \cdot 2(\alpha+\beta) = \frac{\alpha+\beta}{\sigma_q^2}$$

Thus, we have:

$$\frac{\partial}{\partial \beta} D_{\mathrm{KL}}(\hat{p}(\theta_i \mid c) \,\|\, q(\theta_i \mid c)) = \frac{\alpha+\beta}{\sigma_q^2}$$

Hence, we have derived the desired formula for the derivative of the KLD with respect to β.

□

References

N. Choudhary, N. Rao, S. Katariya, K. Subbian, and C. Reddy. Probabilistic entity representation model for reasoning over knowledge graphs. *Advances in neural information processing systems*, 34: 23440–23451, 2021.