4.1 Global Optimality of $p_g = p_{\text{data}}$

We first consider the optimal discriminator D for any given generator G.

Proposition 1. *For G fixed, the optimal discriminator D is*

$$D^*_G(\boldsymbol{x}) = \frac{p_{data}(\boldsymbol{x})}{p_{data}(\boldsymbol{x}) + p_g(\boldsymbol{x})} \tag{2}$$

Proof. The training criterion for the discriminator D, given any generator G, is to maximize the quantity $V(G, D)$

$$\begin{aligned} V(G,D) &= \int_{\boldsymbol{x}} p_{\text{data}}(\boldsymbol{x}) \log(D(\boldsymbol{x}))dx + \int_z p_{\boldsymbol{z}}(\boldsymbol{z}) \log(1 - D(g(\boldsymbol{z})))dz \\ &= \int_{\boldsymbol{x}} p_{\text{data}}(\boldsymbol{x}) \log(D(\boldsymbol{x})) + p_g(\boldsymbol{x}) \log(1 - D(\boldsymbol{x}))dx \end{aligned} \tag{3}$$

For any $(a, b) \in \mathbb{R}^2 \setminus \{0, 0\}$, the function $y \to a \log(y) + b \log(1 - y)$ achieves its maximum in $[0, 1]$ at $\frac{a}{a+b}$. The discriminator does not need to be defined outside of $Supp(p_{\text{data}}) \cup Supp(p_g)$, concluding the proof. □

Theorem 1. *The global minimum of the virtual training criterion $C(G)$ is achieved if and only if $\boxed{p_g = p_{data}}$. At that point, $C(G)$ achieves the value $-\log 4$.*

Proof. For $p_g = p_{\text{data}}$, $D^*_G(\boldsymbol{x}) = \frac{1}{2}$, (consider Eq. 2). Hence, by inspecting Eq. 4 at $D^*_G(\boldsymbol{x}) = \frac{1}{2}$, we find $C(G) = \log \frac{1}{2} + \log \frac{1}{2} = -\log 4$. To see that this is the best possible value of $C(G)$, reached only for $p_g = p_{\text{data}}$, observe that

$$\mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}} [-\log 2] + \mathbb{E}_{\boldsymbol{x} \sim p_g} [-\log 2] = -\log 4$$

define criterion:

$$\begin{aligned} C(G) &= \max_D V(G, D) \\ &= \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}}[\log D^*_G(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}}[\log(1 - D^*_G(G(\boldsymbol{z})))] \\ &= \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}}[\log D^*_G(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{x} \sim p_g}[\log(1 - D^*_G(\boldsymbol{x}))] \\ &= \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}}\left[\log \frac{p_{\text{data}}(\boldsymbol{x})}{P_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}\right] + \mathbb{E}_{\boldsymbol{x} \sim p_g}\left[\log \frac{p_g(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}\right] \end{aligned} \tag{4}$$

$$= \int_x p_{data}(x) \cdot \left[\log\left(\frac{p_{data}(x)}{\frac{p_{data}(x) + p_g(x)}{2}}\right) - \log(2)\right] dx + \int_x p_g(x) \cdot \left[\log\left(\frac{p_g(x)}{\frac{p_{data}(x) + p_g(x)}{2}}\right) - \log(2)\right] dx$$

$$= -\int_x (p_{data}(x) + p_g(x)) \cdot \log(2) + \int_x p_{data}(x) \cdot \log\left(\frac{p_{data}(x)}{\frac{p_{data}(x) + p_g(x)}{2}}\right) dx + \int_x p_g(x) \cdot \log\left(\frac{p_g(x)}{\frac{p_{data}(x) + p_g(x)}{2}}\right) dx$$

$$= -2\log(2) + KL\left(p_{data} \,\Big\|\, \frac{p_{data} + p_g}{2}\right) + KL\left(p_g \,\Big\|\, \frac{p_{data} + p_g}{2}\right)$$

$$\therefore \quad C(G) = -\log(4) + 2 \cdot JSD(p_{data} \,\|\, p_g)$$

b/c $p_{data} = p_g$, $JSD(p_{data} \,\|\, p_g) \geq 0$

Thus, $C^*(G) = -\log(4)$ is optimal when $p_{data} = p_g$