A. Proof of Proposition 1

Proof. Let $\alpha = 1 - p$. If $\mathcal{B}$ happens, then $F_\theta(\mathbf{x}')_y \leq \alpha$ for some $\mathbf{x}' \in \mathbf{x} + \mathcal{S}$, and $\kappa(\theta, \mathbf{x}, y) \geq \tau_L(\alpha)$, therefore[2],

$$\begin{aligned}\Pr_{(\mathbf{x},y)\sim\mathcal{D}}[\mathcal{B}] \leq & \Pr_{(\mathbf{x},y)\sim\mathcal{D}}\left[\kappa(\theta,\mathbf{x},y) \geq \tau_L(\alpha)\right] \\ \leq & \frac{\mathbb{E}[\kappa(\theta,\mathbf{x},y)]}{\tau_L(\alpha)} \qquad \text{(Markov's inequality)} \\ \leq & \frac{\varepsilon}{\tau_L(\alpha)}.\end{aligned}$$

The proof is complete. □

B. Proof of Proposition 2

Proof. Let $\mathcal{B}$ be the event $\{(\exists y' \neq y, \mathbf{x}' \in \mathbf{x} + \mathcal{S})\ C_{F_\theta}(\mathbf{x}') = y'\}$. If $\mathcal{B}$ happens then $F_\theta(\mathbf{x}')_y \leq \frac{1}{2}$ (otherwise $\mathbf{x}'$ will be classified as y), and so $\kappa(\theta, \mathbf{x}, y) \geq \tau_L(1/2)$. On the other hand, if $\mathcal{B}$ does not happen, then we can lower bound $\kappa(\theta, \mathbf{x}, y)$ by 0. Therefore $\varepsilon \geq \mathbb{E}[\kappa(\theta, \mathbf{x}, y)] \geq \Pr[\neg\mathcal{B}] \cdot 0 + \Pr[\mathcal{B}] \cdot \tau_L(1/2) = \Pr[\mathcal{B}] \cdot \tau_L(1/2)$. Tightness follows as we can force equality for each of the inequalities. The proof is complete. □

C. Proof of Proposition 3

Proof. By contraposition it suffices to prove the following

$$\begin{aligned}&\Pr_{(\mathbf{x},y)\sim\mathcal{D}}\left[(\forall y' \neq y, \mathbf{x}' \in N(\mathbf{x},\eta)), \Gamma_\xi^{\mathtt{MCN}}(\mathbf{x}')_{y'} < 1 - p\right] \\ &\geq 1 - q.\end{aligned}$$

By assumption that F satisfies (p, q, δ)-separation, with probability at least $1 - q$ that $(\mathbf{x}, y) \sim \mathcal{D}$, $(\mathbf{x}, y)$ is (p, δ)-good. For every such (p, δ)-good point $(\mathbf{x}, y)$, by assumption, every $\mathbf{z} \in N(\mathbf{x}, \eta)$ is $(p, \mathtt{MCN}_\xi)$-good. Therefore for every such $\mathbf{z}$, $\Gamma_\xi^{\mathtt{MCN}}(\mathbf{z})_y \geq p$, and so $(\forall y' \neq y), \Gamma_\xi^{\mathtt{MCN}}(\mathbf{z})_{y'} < 1 - p$. The proof is complete. □

D. Bounding the probability for (p, q, δ)-separation

This section gives details of our estimation of (p, q, δ)-separation from statistics in Table 1. Note that event $\mathcal{E}_b$ corresponds to a Bernoulli trial. Let $X_1, \ldots, X_t$ be independent indicator random variables, where

$$X_i = \begin{cases} 1 & \text{if } \mathcal{E}_b \text{ happens,} \\ 0 & \text{otherwise} \end{cases},$$

and $X = (\sum_{i=1}^t X_i)/t$. Recall Chebyshev's inequality:

Theorem 1 (Chebyshev's Inequality)**.** *For independent random variables $X_1, \ldots, X_t$ bounded in $[0, 1]$, and $X = (\sum_{i=1}^t X_i)/t$, we have $\Pr[|X - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\mathrm{Var}[X]}{\varepsilon^2}$.*

In our case, $\mathbb{E}[X] = \mathbb{E}[X_1] = \cdots = \mathbb{E}[X_t]$ and let it be μ, and let the computed frequency be $\hat{\mu}$ (observed value). Thus $\Pr[|\hat{\mu} - \mu| \geq \varepsilon] \leq 1/(4\varepsilon^2 t)$ since $\mathrm{Var}[X] = \mu(1 - \mu)/t < 1/4t$. We thus have the following proposition about (p, q, δ)-separation.

Proposition 5. *Let $\alpha, \varepsilon \in [0, 1]$. For sufficiently large t where $\frac{1}{4\varepsilon^2 t} \leq 1 - \alpha$ holds, we have:*

- (Upper bound) *With probability at least α, μ is smaller than $\hat{\mu} + \varepsilon$.*
- (Lower bound) *With probability at least α, μ is bigger than $\hat{\mu} - \varepsilon$.*

For example, we have guarantees for $\alpha = .9$ by putting $\varepsilon = .1$ and $t \geq 250$.

[2] Let X be a nonnegative random variable and $a > 0$, Markov's inequality says that $\Pr[X \geq a] \leq \mathbb{E}[X]/a$.