Conditional Generative Models are No Worse than Unconditional Models: Rigorous Theorem

Setup

Let $(\mathcal{X}, \mathcal{F}_X)$ be the measurable space of data, and $(\mathcal{C}, \mathcal{F}_C)$ be the measurable space of condition variables.

Let $\mu_{X,C}$ be a probability measure on $(\mathcal{X} \times \mathcal{C}, \sigma(\mathcal{X} \times \mathcal{C}))$ representing the joint distribution of data X and condition C.

Let μ_C denote the marginal measure of C:

$$\mu_C(A) = \mu_{X,C}(\mathcal{X} \times A), \quad \forall A \in \mathcal{F}_C.$$

Assume that for μ_C-almost every $c \in \mathcal{C}$, there exists a regular conditional probability measure $\mu_{X|C=c}$ satisfying

$$\mu_{X,C}(B \times A) = \int_A \mu_{X|C=c}(B)\, \mu_C(dc), \quad \forall B \in \mathcal{F}_X, \forall A \in \mathcal{F}_C.$$

Let $\mathcal{P}_{\text{unconditional}}$ be a set of probability measures on $(\mathcal{X}, \mathcal{F}_X)$ representing the unconditional generative model family.

Define the conditional model family

$$\mathcal{P}_{\text{conditional}} = \Big\{ \nu_{X|C} : \forall c \in \mathcal{C},\ \nu_{X|C}(\cdot|c) \in \mathcal{P}_{\text{unconditional}} \Big\}.$$

Let $D(\cdot\|\cdot)$ be a statistical divergence defined on probability measures, satisfying:

1. $D(\mu\|\nu) \geq 0$ with equality if and only if $\mu = \nu$ ν-almost everywhere.

2. D is convex in its first argument: for any $\lambda \in [0,1]$ and measures $\mu_1, \mu_2 \ll \nu$,

$$D(\lambda\mu_1 + (1-\lambda)\mu_2 \| \nu) \leq \lambda D(\mu_1 \| \nu) + (1-\lambda) D(\mu_2 \| \nu).$$

3. D is measurable in the first argument and integrable with respect to μ_C.

Assumption 1 (Conditional distributions are easier to approximate)**.** *For μ_C-almost every $c \in \mathcal{C}$,*

$$\inf_{\nu_X \in \mathcal{P}_{\text{unconditional}}} D\big(\mu_{X|C=c} \,\|\, \nu_X\big) \leq \inf_{\nu_X \in \mathcal{P}_{\text{unconditional}}} D\big(\mu_X \,\|\, \nu_X\big),$$

where the marginal measure μ_X is

$$\mu_X(B) = \int_{\mathcal{C}} \mu_{X|C=c}(B)\, \mu_C(dc), \quad \forall B \in \mathcal{F}_X.$$

Assumption 2 (Independent conditional choice)**.** *For each $c \in \mathcal{C}$, there exists an optimal unconditional measure*

$$\nu^*_{X|C=c} = \arg \inf_{\nu_X \in \mathcal{P}_{\text{unconditional}}} D(\mu_{X|C=c} \| \nu_X).$$

Assumption 3 (Measurability and integrability)**.** *The map $c \mapsto \mu_{X|C=c}$ is measurable, and*

$$\int_{\mathcal{C}} D\big(\mu_{X|C=c} \,\|\, \nu_{X|C=c}\big)\, \mu_C(dc) < \infty.$$

Theorem (Conditional models no worse than unconditional)

Under the above assumptions, the optimal conditional model achieves an average divergence no larger than the optimal unconditional model:

$$\inf_{\nu_{X|C} \in \mathcal{P}_{\text{conditional}}} \int_{\mathcal{C}} D\big(\mu_{X|C=c} \,\|\, \nu_{X|C=c}\big)\, \mu_C(dc) \leq \inf_{\nu_X \in \mathcal{P}_{\text{unconditional}}} D\big(\mu_X \,\|\, \nu_X\big).$$

If Assumption 1 is strict for a set of positive μ_C measure, the inequality is strict.