Model Evaluation Framework

1 OT-based stability evaluation criterion

1.1 Definition 1 (OT discrepancy with moment constraints)

If:

$\mathcal{Z} \subseteq \mathbb{R}^d, \mathcal{W} \subseteq \mathbb{R}_+$: convex, closed sets,

$c : (\mathcal{Z} \times \mathcal{W})^2 \to \mathbb{R}_+$: lower semicontinuous function,

$\mathbb{Q}, \mathbb{P} \in \mathcal{P}(\mathcal{Z} \times \mathcal{W})$

Then:

$\mathrm{M}_c : \mathcal{P}(\mathcal{Z} \times \mathcal{W})^2 \to \mathbb{R}_+$ is a function, defined through

$$\mathrm{M}_c(\mathbb{Q}, \mathbb{P}) = \begin{cases} \inf & \mathbb{E}_\pi[c((Z, W), (\hat{Z}, \hat{W}))] \\ \text{s.t.} & \pi \in \mathcal{P}\left((\mathcal{Z} \times \mathcal{W})^2\right) \\ & \pi_{(Z,W)} = \mathbb{Q}, \pi_{(\hat{Z},\hat{W})} = \mathbb{P} \\ & \mathbb{E}_\pi[W] = 1 \quad \pi\text{-a.s} \end{cases}$$

is called OT discrepancy wth moment constraints induced by $c, \mathbb{Q}, \mathbb{P}$

f_β : given learning model, trained on the distribution:

$\mathbb{P}_0 \in \mathcal{P}(\mathcal{Z})$, we have:

Problem P

$$\mathfrak{R}(\beta, r) = \begin{cases} \inf_{\mathbb{Q} \in \mathcal{P}(\mathcal{Z} \times \mathcal{W})} & \mathrm{M}_c(\mathbb{Q}, \hat{\mathbb{P}}) \\ \text{s.t.} & \mathbb{E}_\mathbb{Q}[W \cdot \ell(\beta, Z)] \geq r \end{cases}$$

- $\hat{\mathbb{P}}$ is selected as $\mathbb{P}_0 \otimes \delta_1$
  - δ_1 : Dirac delta function
  - $\mathrm{M}_c(\mathbb{Q}, \hat{\mathbb{P}})$: OT discrepancy with moment constraints between the projected distribution $\mathbb{Q}$ and the reference distribution $\hat{\mathbb{P}}$
  - $l(\beta, z)$: prediction risk of model f_β on sample z
  - $r > 0$: pre-defined risk threshold

"The best way to transfer probability distribution from A to B"

z: data point

w: weight

π: policy

Example c:

Formula 1

$$c((z, w), (\hat{z}, \hat{w})) = \theta_1 \cdot w \cdot d(z, \hat{z}) + \theta_2 \cdot (\phi(w) - \phi(\hat{w}))_+$$

- $d(z, \hat{z}) = \|x - \hat{x}\|_2^2 + \infty \cdot |y - \hat{y}|$: cost with different $z, \hat{z}$
- $(\phi(w) - \phi(\hat{w}))_+$: cost related to differences in probability mass.
  - $\phi : \mathbb{R}_+ \to \mathbb{R}_+$: convex function, where:
  - $\phi(1) = 0$
- $\theta_1, \theta_2 \geq 0$: hyperparameters, where:
  - $\dfrac{1}{\theta_1} + \dfrac{1}{\theta_2} = C$ for some constant C

The minimum deviation needed to make this model risky

$\mathfrak{R}$: risk

Dual reformulation and its interpretation

Theorem 1 (Strong duality

Skip the proof

for problem for problem (P))

Suppose:

- $\mathcal{Z} \times \mathcal{W}$ is compact.
- $l(\beta, \cdot)$ is upper semicontinuous for all β
- $c : (\mathcal{Z} \times \mathcal{W})^2 \to \mathbb{R}_+$ is continuous
- $r < \bar{r} := \max_{z \in \mathcal{Z}} l(\beta, z)$

Then:

Function D

$\mathfrak{R}(\beta, r) = \sup_{h \in \mathbb{R}_+, \alpha \in \mathbb{R}} hr + \alpha + \mathbb{E}_{\hat{\mathbb{P}}} \left[\tilde{\ell}_c^{\alpha,h}(\beta, (\hat{Z}, \hat{W})) \right]$

- $\tilde{\ell}_c^{\alpha,h}(\beta, (\hat{Z}, \hat{W}))$: surrogate function
  - it equals to : $\min_{(z,w) \in \mathcal{Z} \times \mathcal{W}} c((z, w), (\hat{z}, \hat{w})) + \alpha w - h \cdot w \cdot l(\beta, z)$, for all $\hat{z} \in \mathcal{Z}, \hat{w} \in \mathcal{W}$.

(is it $+aw$? in the proof it's $-aw$)

Proof for Function D

Reformulate Problem (P) into a infinite-dimension linear program:

Formula Primal

$$\begin{array}{ll} \inf_\pi & \mathbb{E}_\pi[c((Z, W), (\hat{Z}, \hat{W}))] \\ \text{s.t.} & \pi \in \mathcal{P}\left((\mathcal{Z} \times \mathcal{W})^2\right) \\ & r - \mathbb{E}_\pi[W \cdot \ell(\beta, Z)] \leq 0 \\ & \mathbb{E}_\pi[W] = 1 \\ & \pi_{(\hat{Z}, \hat{W})} = \hat{\mathbb{P}}. \end{array}$$

We get the Lagrangian function

$L(\pi; h, \alpha) = hr + \alpha + \mathbb{E}_\pi[c((Z, W), (\hat{Z}, \hat{W})) - h \cdot W \cdot \ell(\beta, Z) - \alpha \cdot W]$,

where $h \in \mathbb{R}_+, \alpha \in \mathbb{R}, \pi$ belongs to :

- $\Pi_{\hat{\mathbb{P}}} = \left\{ \pi \in \mathcal{P}((\mathcal{Z} \times \mathcal{W})^2) : \pi_{(\hat{Z}, \hat{W})} = \hat{\mathbb{P}} \right\}$

$\mathcal{Z} \times W$ is compact

$\Rightarrow \mathcal{P}(\mathcal{Z} \times \mathcal{W})$ is tight.

$\Rightarrow \Pi_{\hat{\mathbb{P}}}$ is tight

$\Rightarrow \Pi_{\hat{\mathbb{P}}}$ has a compact closure (Prokhorov's theorem)

$\Pi_{\hat{\mathbb{P}}}$ is weakly closed

$\Rightarrow \Pi_{\hat{\mathbb{P}}}$ is compact (tight + close)

$\Pi_{\hat{\mathbb{P}}}$ is convex

Prove $L(\pi; h, \alpha)$ is lower semicontinuous in π under the weak topology

Suppose:

π_n converges weakly to π

$\Rightarrow \liminf_{n \to +\infty} \int g \mathrm{d}\pi_n \geq \int g \mathrm{d}\pi$, for any lower semicontinuous function g that is bounded below (Portmanteau theorem)

$l(\beta, \cdot)$ is upper semicontinuous for all β,

and $w, h \geq 0$,

$\Rightarrow h \cdot w \cdot l(\beta, z)$ is upper semicontinuous, w.r.t (z, w)

$c((z, w), (\hat{z}, \hat{w}))$ is lower semicontinuous

$\Rightarrow c((z, w), (\hat{z}, \hat{w})) - h \cdot w \cdot l(\beta, z) - \alpha \cdot w$ is lower semicontinuous w.r.t (z, w) for any $(\hat{z}, \hat{w}) \in \mathcal{Z} \times \mathcal{W}$

$\mathcal{Z} \times \mathcal{W}$ is compact

$\Rightarrow$ the function is bounded below

$\Rightarrow \liminf_{n \to +\infty} L(\pi_n; h, \alpha) \geq L(\pi; h, \alpha)$

$\Rightarrow L(\pi; h, \alpha)$ is lower semicontinuous in π under the weak topology

Prove continuous in (h, α) under the uniform topology in $\mathbb{R}_+ \times \mathbb{R}$

Suppose:

$\lim_{n\to+\infty} h_n = h$ in Euclidean topology, $\lim_{n\to\infty} |\alpha_n| < \bar{\alpha}$ in Euclidean topology

Exists:

$\bar{h} \in \mathbb{R}_+, \bar{\alpha} \in \mathbb{R}$, with $\sum_{n\to\infty} |h_n| \le \bar{h}, \sup_{n\to\infty} |\alpha_n| < \bar{\alpha}$, for all $n \ge 1$

$\Rightarrow \lim_{n\to+\infty} L(\pi; h_n, \alpha_n) = L(\pi; h, \alpha)$ (dominated convergence theorem)

$\Rightarrow L(\pi; h, \alpha)$ is continuous in (h, α) under the Ecludiean topology in $\mathbb{R}_+ \times \mathbb{R}$

Formula 5

$\Rightarrow \inf_{\pi\in\Pi_{\hat{\mathbb{P}}}} \sup_{h\in\mathbb{R}_+,\alpha\in\mathbb{R}} L(\pi; h, \alpha) = \sup_{h\in\mathbb{R}_+,\alpha\in\mathbb{R}} \inf_{\pi\in\Pi_{\hat{\mathbb{P}}}} L(\pi; h, \alpha)$ (Sion's minmax theorem)

Rewrite:

$L(\pi; h, \alpha) = \mathbb{E}_\pi\left[c((Z, W), (\hat{Z}, \hat{W}))\right] + h\left(r - \mathbb{E}_\pi[W \cdot \ell(\beta, Z)]\right) + \alpha\left(1 - \mathbb{E}_\pi[W]\right)$ (The original paper lost a close bracket)

$\Rightarrow \inf_{\pi\in\Pi_{\hat{\mathbb{P}}}} \sup_{h\in\mathbb{R}_+,\alpha\in\mathbb{R}} L(\pi; h, \alpha)$ is bounded above

We construct:

$\mathbb{Q}_0 = \delta_{(z^*,1)}$

- $z^* = \arg\max_{z\in\mathcal{Z}} l(\beta, z)$

Then:

$$\begin{aligned}
&\inf_{\pi\in\Pi_{\mathbb{P}}} \sup_{h\in\mathbb{R}_+,\alpha\in\mathbb{R}} L(\pi; h, \alpha) \\
&\le \sup_{h\in\mathbb{R}_+,\alpha\in\mathbb{R}} L\left(\mathbb{Q}_0 \otimes \hat{\mathbb{P}}; h, \alpha\right) \\
&= \mathbb{E}_{\mathbb{Q}_0\otimes\hat{\mathbb{P}}}[c((Z, W), (\hat{Z}, \hat{W}))] + \sup_{h\in\mathbb{R}_+} h(r - \bar{r}) \quad (E_\pi[W] = 1) \\
&< +\infty
\end{aligned}$$

- $\bar{r} = \mathbb{E}_{\mathbb{Q}_0}[l(\beta, Z)]$ (Notice W is independent with it) $= \max_{z\in Z} l(\beta, Z)$
- combined c is continuous, it's bounded on a compact domain $Z \times W$ (The suppose)

$$\begin{aligned}
\Rightarrow & r - \mathbb{E}_\pi[W \cdot l(\beta, Z)] \le 0 \\
& \mathbb{E}_\pi[W] = 1
\end{aligned}$$

Then: for (5) right hand side:

$$\sup_{h\in\mathbb{R}_+,\alpha\in\mathbb{R}} \inf_{\pi\in\Pi_{\mathrm{P}}} L(\pi;h,\alpha).$$

$$= \sup_{h\in\mathbb{R}_+,\alpha\in\mathbb{R}} hr+\alpha+\inf_{\pi\in\Pi_{\mathbb{R}}} \mathbb{E}_{\pi}[c((Z,W),(\hat{Z},\hat{W}))-h\cdot W\cdot \ell(\beta,Z)-\alpha\cdot W].$$

Notice:

$$\inf_{\pi\in\Pi_{\mathfrak{R}}} \mathbb{E}_{\pi}[c((Z,W),(\hat{Z},\hat{W}))-h\cdot W\cdot \ell(\beta,Z)-\alpha\cdot W]$$
$$=\mathbb{E}_{\hat{\mathbb{P}}}\left[\min_{(z,w)\in\mathcal{Z}\times\mathcal{W}} c((z,w),(\hat{Z},\hat{W}))-h\cdot w\cdot \ell(\beta,z)-\alpha\cdot w\right],$$

End Proof for Function D

Proposition 1 (Dual reformulations)

Suppose:

$\mathcal{W}=\mathbb{R}_+$

(i) If:

$\phi(t)=t\log t-t+1$, (D) admits:

function 2

$\sup_{h\geq 0} hr-\theta_2\log\mathbb{E}_{\mathbb{P}_0}\left[\exp\left(\frac{l_{h,\theta_1}(\hat{Z})}{\theta_2}\right)\right]$

(ii) If:

$\phi(t)=(t-1)^2$, (D) admits:

function 3

$\sup_{h\geq 0,\alpha\in\mathbb{R}} hr+\alpha+\theta_2-\theta_2\mathbb{E}_{\mathbb{P}_0}\left[\left(\frac{\ell_{h,\theta_1}(\hat{Z})+\alpha}{2\theta_2}+1\right)_+^2\right]$

- $l_{h,\theta_1}(\hat{z}):=\max_{z\in\mathcal{Z}} h\cdot l(\beta,z)-\theta_1\cdot d(z,\hat{z})$: the d-trasform of $h\cdot l(\beta,\cdot)$ with the step size θ_1