Scoring Rules

Lecture 6

GOALS //

- Objective Functions For Probabilistic Estimation
- Scoring Rules ch 11
- Learning Filters From Scoring Rules. ch 4
- Amortization ch 10

$\boxed{\text{WORK UNDER Data Assumption } \circledast \text{, Lecture 4}}$

OBJECTIVES (PROBABILISTIC ESTIMATION) //

- objectives so far use distance in $\mathbb{R}^d$.
- hence address state estimation.
- Instead try objectives in $P(\mathbb{R}^d)$.

For EnKF define

$$\mu_j^N(\theta) = \frac{1}{N} \sum_{n=1}^{N} \delta_{v_j^{(n)}(\theta)}$$

Recall μ_j true filter & define

$$J(\theta) = \frac{1}{J} \sum_{j=1}^{J} D(\mu_j^N(\theta) \,\|\, \mu_j)$$

Problem: we don't have μ_j in useable form.

SCORING RULES //

Definition/ $S: P(\mathbb{R}^d) \times \mathbb{R}^d$ is a scoring rule

The expected score is $\bar{S}: P(\mathbb{R}^d) \times P(\mathbb{R}^d) \to \mathbb{R}$ defined by $\bar{S}(\rho_1, \rho_2) = \mathbb{E}^{v \sim \rho_2} S(\rho_1, v)$ □

When does expected score define a divergence?

Definition/ A scoring rule is proper if, for all $\rho_1, \rho_2 \in \mathbb{R}^d$, $\bar{S}(\rho_2, \rho_2) \leq \bar{S}(\rho_1, \rho_2)$. It is strictly proper if equality holds iff $\rho_1 = \rho_2$.

Theorem/ If S is a strictly proper scoring rule then

$$D_S(\rho_1 \| \rho_2) = \bar{S}(\rho_1, \rho_2) - \bar{S}(\rho_2, \rho_2)$$

is a divergence.

Example/ $ES: P(\mathbb{R}^d) \times \mathbb{R}^d$ is defined by

$$ES(\rho, v) = \mathbb{E}^{u \sim \rho} |u - v| - \frac{1}{2} \mathbb{E}^{(u, u') \sim \rho \otimes \rho} |u - u'|.$$

Then $\overline{ES}(\rho_1,\rho_2) = \mathbb{E}^{v\sim\rho_2} ES(\rho_1, v)$.

Lemma/

$$\frac{1}{2} D_E^2(\rho_1,\rho_2) = D_{ES}(\rho_1 \| \rho_2)$$

$$:= \overline{ES}(\rho_1,\rho_2) - \overline{ES}(\rho_2,\rho_2).$$

Example/ Let $d=1$ & let F_ρ denote the cdf of ρ. The continuous ranked probability score is $CRPS: P(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ defined by

$$CRPS(\rho, v) = \int_{\mathbb{R}^d} \left(F_\rho(u) - \mathbb{I}_{u \geq v}(u)\right)^2 du.$$

Note: $\mathbb{I}_{u\geq v}(u)$ is cdf of δ_v.

Lemma/ $CRPS(\rho, v) = ES(\rho, v)$ in $d=1$.

LEARNING FILTERS FROM SCORING RULE//

Recall the unimplementable

$$J(\theta) = \frac{1}{J} \sum_{j=1}^{J} D(\mu_j^N(\theta) \| \mu_j)$$

Replace this by

$$J(\theta) = \frac{1}{J} \sum_{j=1}^{J} S(\mu_j^n(\theta), v_j^+)$$

AMORTIZED FILTERING //

Data Assumption / we are given

$$\{ (v_j^+)^{(n)}, (y_j^+)^{(n)} \}_{j \in \{1, \dots, J\}}^{n \in \{1, \dots, N\}}$$

Theorem / Let $S : \mathcal{P}(\mathbb{R}^{d_v}) \times \mathbb{R}^{d_v} \to \mathbb{R}$ be strictly proper. Then

$$\mathbb{E}^{(v_j^+, y_j^+)}\left[S(q, y_j^+) \right] \geq \mathbb{E}^{(v_j^+, y_j^+)}\left[S(\pi_j, y_j^+) \right]$$

$\square$

Suggests:

$$J(\theta) = \frac{1}{J} \sum_{j=1}^{J} \mathbb{E}^{(v_j^+, y_j^+)}\left[S(\mu_j^n(\theta), y_j^+) \right]$$

$$\theta^* = \arg\inf_{\theta} J(\theta)$$