Universal coding, intrinsic volumes, and metric complexity

Sequential probability assignment

Data sequence $y_{1:n} = (y_1, \dots, y_n) \in \mathcal{Y}^n$ (unknown).
Goal: assign probabilities to possible sequences $y_{1:n}$
Log-loss/log-likelihood: for proba. density p on $\mathcal{Y}^n$,

$$\ell(p, y_{1:n}) = -\log p(y_1, \dots, y_n).$$

Coding interpretation: proba. densities p on $\mathcal{Y}^n$ (for finite $\mathcal{Y}$) $\overset{\sim}{\longleftrightarrow}$ "codes" $\phi : \mathcal{Y}^n \to \bigcup_{k \geqslant 1} \{0,1\}^k$, such that $\mathrm{length}(\phi(y_{1:n})) \simeq -\log_2 p(y_{1:n}) \pm 1$.
Sequentially: after seeing $y_{1:i-1}$, predict y_i by $p_i(\cdot | y_{1:i-1})$

$$p(y_{1:n}) = \prod_{i=1}^n p_i(y_i | y_{1:i-1}), \quad \sum_{i=1}^n \log p_i(y_i | y_{1:i-1}) = \log p(y_{1:n}).$$

Minimax regret

Statistical model on data: set $\mathcal{P}$ of densities p on $\mathcal{Y}^n$.
Statistical setting: random sequence $Y_{1:n} \sim p^* \in \mathcal{P}$. For p a density, $\mathbf{E}[\ell(p, Y_{1:n}) - \ell(p^*, Y_{1:n})] = \mathrm{KL}(p^*, p)$.
Deterministic setting: no assumption on sequence $y_{1:n} \in \mathcal{Y}^n$, but use model $\mathcal{P}$ as a benchmark/oracle:

$$\mathrm{Regret}(q, \mathcal{P}, y_{1:n}) = \ell(q, y_{1:n}) - \inf_{p \in \mathcal{P}} \ell(p, y_{1:n}).$$

Complexity measure: minimax regret *(Shtarkov '87)*

$$\mathrm{R}^*(\mathcal{P}) = \inf_q \sup_{y_{1:n} \in \mathcal{Y}^n} \mathrm{Regret}(q, \mathcal{P}, y_{1:n}).$$

Gaussian setting

Real observations: $\mathcal{Y} = \mathbf{R}$. Gaussian noise: model $\mathcal{P}_A = \{\mathcal{N}(\theta, I_n) : \theta \in A\}$ for some constraint set $A \subset \mathbf{R}^n$.
Regression: observations $y_i = \theta_i + \varepsilon_i$, $\theta \in A$, $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0,1)$.

Intrinsic volumes of a convex body

$K \subset \mathbf{R}^n$ **convex body**, $1 \leqslant j \leqslant n$, j-th intrinsic volume:

$$V_j(K) = c_{n,j} \, \mathbf{E} \, \mathrm{Vol}_j(P_{\mathrm{F}} K), \quad P_{\mathrm{F}} : \text{orth. proj. on F},$$

with F random subspace of $\mathbf{R}^n$ of dim. j.

- $V_1(K/\sqrt{2\pi}) = w(K) = \mathbf{E}_{X \sim \mathcal{N}(0, I_n)} \sup_{\theta \in K} \langle \theta, X \rangle$ **Gaussian width**; $V_{n-1}(K) = \frac{1}{2} \mathrm{surface}(\partial K)$, $V_n = \mathrm{Vol}_n$.

Steiner's formula: for $r \geqslant 0$, letting B_2^j the ℓ_2 ball in $\mathbf{R}^j$,

$$\mathrm{Vol}_n(K + r B_2^n) = \sum_{j=0}^n V_{n-j}(K) \kappa_j r^j, \quad \kappa_j = \mathrm{Vol}_j(B_2^j).$$

Minimax regret and intrinsic volumes

Model: $\mathcal{P}_A = \{\mathcal{N}(\theta, I_n) : \theta \in A\}$ for $A \subset \mathbf{R}^n$.
For any convex body $K \subset \mathbf{R}^n$,

$$\mathrm{R}^*(K) = \log \left(\sum_{j=0}^n V_j\Big(\frac{K}{\sqrt{2\pi}}\Big) \right) = \log W(K/\sqrt{2\pi}). \qquad (1)$$

Connects universal coding to convex geometry.
$W = \sum_{j=0}^n V_j$ called Wills functional in convexity.
For general A, one has $\mathrm{R}^*(A) = \log W(A/\sqrt{2\pi})$, defining

$$W(A) = \int_{\mathbf{R}^n} e^{-\pi \mathrm{dist}^2(x, A)} \mathrm{d}x.$$

Comparison inequality

Key property: $\mathrm{R}^*(A)$ is a "metric" quantity.
If $\varphi : A \to \mathbf{R}^n$ contraction (1-Lip.), $\mathrm{R}^*(\varphi(A)) \leqslant \mathrm{R}^*(A)$.
Analogous to **Slepian-Sudakov-Fernique** comparison theorem for the Gaussian width: $w(\varphi(A)) \leqslant w(A)$.
True for **general** (nonconvex) $A \subset \mathbf{R}^n$.

Isomorphic characterization

Alternative parameters: local Gaussian width

$$\sup_{\theta \in A} w\big(A \cap B(\theta, r)\big)$$

Global covering numbers: $N(A, r)$ = number of balls of radius r needed to cover A. For any $A \subset \mathbf{R}^n$,

$$\mathrm{R}^*(A) \asymp \inf_{r \geqslant 0} \Big\{ \sup_{\theta \in A} w\big(A \cap B(\theta, r)\big) + \log N(A, r) \Big\}. \qquad (2)$$

Implies characterization of the log-Laplace transform of intrinsic volumes of a convex body K: for $\lambda \geqslant 0$,

$$\log \Big(\sum_{j \geqslant 0} V_j(K) \lambda^j \Big) \asymp \inf_{r \geqslant 0} \big\{ \lambda w\big((K-K) \cap r B_2\big) + \log N(K, r) \big\}.$$

Metric estimates

Goal: explicit "metric" expression. For the Gaussian width: Talagrand's majorizing measures theorem

$$w(A) \asymp \inf_{(\mathcal{A}_j)_{j \geqslant 0}} \sup_{\theta \in A} \sum_{j \geqslant 0} 2^{j/2} \operatorname{diam} A_j(\theta), \qquad |\mathcal{A}_j| \leqslant 2^{2^j} - 1;$$

$$\mathrm{R}^*(A) \asymp \inf_{(\mathcal{A}_j)_{j \geqslant 0}} \inf_{p \geqslant 0} \Big\{ 2^p - 1 + \sup_{\theta \in A} \sum_{j \geqslant p} 2^{j/2} \operatorname{diam} A_j(\theta) \Big\}.$$

Additive properties

Log-concavity: $W(\lambda A + (1-\lambda) B) \geqslant W(A)^\lambda W(B)^{1-\lambda}$ (Alonso-Gutiérrez, Hernández Cifre, Yepes Nicolás '21), $\lambda \in [0,1]$.
Sub-multiplicativity: $W(A + B) \leqslant W(A) W(B)$.

Example: ellipsoids

If $E = \{\theta \in \mathbf{R}^n : \sum_{i=1}^n \theta_i^2 / a_i^2 \leqslant 1\}$ ellipsoid
$\mathrm{R}^*(E) \asymp \inf_{r > 0} \big\{ \sum_{i=1}^n \log\big(1 + a_i^2/r^2\big) + r^2 \big\}$.

Jaouad Mourtada
Department of Statistics
CREST, ENSAE Paris