CWI

Wouter Koolen

Optimization for pure exploration

linear   linear   convex quadratic smooth

$$\min_{q} \max_{w} \min_{\lambda} f(q, w, \lambda)$$

$\in \Delta_J$ $\in \Delta_K$ $\mathbb{R}^d$

Best Arm Identification

seq. dec. problem. (drug testing, A/B testing,)

K "arms"

hidden vector $\mu \in \mathbb{R}^K$ mean rewards, Gaussian $\sigma = 1$

Interaction protocol:

for $t = 1, 2, \dots, \tau$

- pick $A_t \in [K]$
- see $X_t \overset{\text{indep}}{\sim} \mathcal{N}(\mu_{A_t}, 1)$

output recommendation $\hat{\imath} \in [K]$

→ confidence ← fix

→ fast ← optimize

look at δ-correct algorithms

$$\forall \mu : \ \mathbb{P}_\mu(\hat{\imath} \neq \underbrace{i^*(\mu)}_{\arg\max_k \mu_k}) \leq \delta$$

$$(A_1, X_1, \dots, A_\tau, X_\tau, \hat{\imath}) =: H$$

the sample complexity at μ is $\mathbb{E}_\mu[\tau]$

Lower bds either at μ or at λ $\quad i^*(\mu) \neq i^*(\lambda)$

$$KL\left(\mathbb{P}_\mu^{alg}(H) \,\|\, \mathbb{P}_\lambda^{alg}(H)\right)$$

$$\geq KL\left(\underbrace{\mathbb{P}_\mu^{alg}(\hat{\imath} = i^*(\mu))}_{\geq 1-\delta} \,\|\, \underbrace{\mathbb{P}_\lambda^{alg}(\hat{\imath} = i^*(\mu))}_{\leq \delta}\right)$$

$$\geq$$

$$\approx \ln 1/\delta$$

$$= \underbrace{\sum_{a=1}^K \underbrace{\mathbb{E}_\mu[N_a(\tau)]}_{\text{does not depend on } \lambda} \underbrace{KL(\mu_a \| \lambda_a)}_{\frac{1}{2}(\mu_a - \lambda_a)^2}}$$

so:

$$\inf_{\lambda : i^*(\lambda) \neq i^*(\mu)} \qquad \geq \ln 1/\delta$$

$$\mathbb{E}_\mu[\tau] \inf_{\lambda \dots} \sum_{a=1}^K \frac{\mathbb{E} N_a(\tau)}{\mathbb{E}\tau} KL(\mu_a, \lambda_a) \geq \ln 1/\delta$$

$$\mathbb{E}_\mu[\tau] \underbrace{\sup_{w \in \Delta_K} \inf_{\lambda \dots} \sum_{a=1}^K w_a KL(\mu_a, \lambda_a)} \geq \ln 1/\delta$$

$$\mathbb{E}_\mu[\tau] \geq \frac{\ln 1/\delta}{f_a(\mu)}$$